SPRINGER BRIEFS IN ELECTRICAL AND COMPUTER ENGINEERING · SIGNAL PROCESSING

Ee-Leng Tan
Woon-Seng Gan

Perceptual Image Coding with Discrete Cosine Transform

Springer

SpringerBriefs in Electrical and Computer Engineering

Signal Processing

More information about this series at http://www.springer.com/series/11560

Ee-Leng Tan · Woon-Seng Gan

Perceptual Image Coding with Discrete Cosine Transform

Springer

Ee-Leng Tan
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore
Singapore

Woon-Seng Gan
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore
Singapore

ISSN 2191-8112 ISSN 2191-8120 (electronic)
SpringerBriefs in Electrical and Computer Engineering
ISSN 2196-4076 ISSN 2196-4084 (electronic)
SpringerBriefs in Signal Processing
ISBN 978-981-287-542-6 ISBN 978-981-287-543-3 (eBook)
DOI 10.1007/978-981-287-543-3

Library of Congress Control Number: 2015939147

Springer Singapore Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer Science+Business Media Singapore Pte Ltd. is part of Springer Science+Business Media (www.springer.com)

Acknowledgments

I would like to express my deepest gratitude to my Ph.D. supervisor, Assoc. Prof. Gan Woon-Seng, for his invaluable guidance and encouragement throughout the course of this work, which led to this monograph. I sincerely thank friends and colleagues, Dr. Lee Kong Aik, Dr. Wang Liang, Dr. Wen Yuan, and Mr. Furi Andi Karnapi, for being encouraging and supportive of my work. Last and not least, I would like to thank all the friends and colleagues, who participated in the subjective testing reported in this monograph, as well as providing me with useful feedback and comments.

Contents

Chapter 1
Introduction

One of the biggest challenges in the field of image and video processing is evaluating and optimizing the quality of digital imaging system with respect to storage capacity and transmission bandwidth of the digital imaging system. Since the ultimate receiver for most imagery systems is human, therefore, it is pertinent to take into account the human visual system (HVS) in such systems to achieve optimal performance in the perceptual sense. It is well-known that the physiological and psychological mechanisms of the HVS prevent it from detecting all changes in an image. By exploiting the limitations of HVS, storage capacity and transmission bandwidth of digital imaging system can be optimally allocated for optimal visual experience.

One typical component of such a digital imaging system would be the perceptual image coder (PIC). Perceptually-tuned image compression improves coding efficiency of images while minimizes the amount of perceptible distortion added into the compressed image. Specifically, changes in the compressed image are undetectable by the HVS if these changes are lower than the just-noticeable-difference (JND) threshold. To date, numerous computational models for JND have been proposed, and these models can be computed from subbands or pixels of an image. These computational models determine the JND of pixels or subbands of an image, and account for three visual factors of the HVS, namely, contrast sensitivity function (CSF), luminance adaptation, and contrast masking.

A survey on classic as well as recent computational models shall be presented in this monograph. We will also review the three visual factors (contrast sensitivity factor, luminance adaptation, and contrast masking) applied in these computational models. Since discrete cosine transform (DCT) is applied in many image and video standards (JPEG, MPEG-1/2/4, H.261/3), we focus our survey on the computational models for JND that are based on DCT. We shall also present a comparative analysis of the computational models using quantitative and qualitative performance evaluation, which compares the noise shaping performance of the computational models with subjective evaluation, and the accuracy between the estimated JND thresholds and subjective evaluation.

E.-L. Tan and W.-S. Gan, *Perceptual Image Coding with Discrete Cosine Transform*, SpringerBriefs in Signal Processing,
DOI 10.1007/978-981-287-543-3_1

1.1 Outline

This monograph begins with an introduction of the cortex filters and the frequency decomposition using cortex filters. The cortex filters are shown to provide good approximation of the multi-channel response of HVS [Wat87]. Next, the widely adopted spatial CSF proposed by Ahumada and Peterson [AP92] is discussed. Subsequently, the base detection threshold derived from the spatial CSF is adjusted by the effects of the local mean luminance and the local spatial content, which are referred as luminance adaptation and contrast masking, respectively. Finally, several techniques estimating the luminance adaptation and contrast masking from subbands and pixels of an image are reviewed.

In Chap. 3, we present a brief introduction of the four modes of operation defined in the joint photographic experts group (JPEG) standard [ISO94]. We show how image coders can be integrated with JND models in the Type-II DCT (DCT-II) domain [Wat93, ZLX05, WN09] and pixel domain [CL95, YLL05]. These coders shall be referred as PICs and are compatible with the JPEG standard. Comparative analysis and discussions of the key modules of these PICs, such as luminance adaptation, block classification, non-linear additive masking model (NAMM) are presented as well.

In Chap. 4, we perform several comparative analysis of JND models that are computed from subbands [Wat93, WN09, ZLX05] and pixels [CL95, YLL05]. Specifically, the noise-shaping performance and the contrast sensitivity of these JND models are compared. In addition, the performance analysis of these JND models is studied by comparing their correlation between these JND models and the human perception of visual degradation.

Finally, our concluding remarks are presented in Chap. 5.

Chapter 2
Computational Models for Just-Noticeable Differences

The growing demand for transmission and storage of images has spurred much effort in improving image compression techniques. To achieve this goal, one promising approach is to integrate properties of the HVS into image compression techniques [JJS93]. The central idea of such approach is to embed coding distortion beneath the spatial visibility threshold of the HVS. This threshold is commonly referred as the JND threshold [JJS93] as it specifies the minimum sensory difference that is detectable by the HVS. In the context of image compression, a perceptually perfect image is obtained at the lowest possible bit-rate [JJS93] if the coding error of each pixel in a compressed image is exactly at level of JND. Over the years, several computational models for JND have been developed and employed in image compression. These computational models for JND models are computed using subbands [SJ89, Wat93, TS96, HK00, HK02, ZLX05, ZLX08, WN09] and pixels [CL95, CC96, CB99, YLL03, YLL05, LLP10] of an image.

The first few models of the HVS [Sch56, MS74, Fau79] were developed using a single channel approach. Such models regard the HVS as a single spatial filter, which is defined by the CSF. One of the first few HVS based image quality metrics for luminance images was developed by Mannos and Sakrison [MS74]. By inferring some properties of the human vision from psychophysical experiments, Mannos and Sakrison derived a closed-form expression describing the contrast sensitivity of the HVS as a function of spatial frequency.

It is later argued that the HVS is a multi-channel system with each channel tuned to different ranges of spatial frequencies and orientations [Dau80], and many multi-channel models were subsequently proposed. Multi-channel HVS models are employed in metrics such as visual differences predictor (VDP) proposed by Daly [Dal92, Dal93], and the visual discrimination model (VDM) proposed by Lubin [Lub93, Lub95]. These image quality metrics are intended for general applicability, but are computationally expensive to implement.

A priori knowledge of the image processing algorithm (such as image compression) permits the use of specialized vision models. Although specialized vision models are not as versatile as the generalized models, specialized models can perform very well in a given application scope. Such vision models are usually simpler and computationally efficient. One example of an image coder based on a

E.-L. Tan and W.-S. Gan, *Perceptual Image Coding with Discrete Cosine Transform*, SpringerBriefs in Signal Processing,
DOI 10.1007/978-981-287-543-3_2

specialized vision model is the DCTune [Wat93], which permits higher image compression by exploiting the limitations of the HVS.

The general block diagram of a computational model for JND is shown in Fig. 2.1. Before the computational model for JND is applied, pre-processing such as color space transformation and frequency decomposition might be performed on the input image. In general, most JND models incorporate four properties of the HVS, namely, spatial contrast sensitivity, luminance adaptation, contrast masking, and temporal masking [Bov05]. The last three properties of the HVS are considered as elevation parameters of the base threshold which are determined by the spatial contrast sensitivity.

Since no masking is present in the measurement of contrast sensitivity, the effect of the background luminance on contrast sensitivity is typically accounted as luminance adaptation, or luminance masking [Wat93]. Contrast masking refers to the change of visibility of one image component due to the presence of another. The strongest contrast masking occurs when both components are of the same or at similar spatial frequency, orientation, and location. Temporal masking refers to the reduced contrast sensitivity due to the temporal variation of light intensity falling into the eye, and is commonly adopted in video compression. Since this monograph focuses on image processing, only the first three properties of the HVS shall be introduced in the following sections of this chapter.

The input image is decomposed into several components (also known as channels or subbands) in multi-channel HVS models. Numerous decomposition methods are used in PICs and image quality metrics, which include Fourier decomposition [CR68, MS74], Gabor decomposition [Dau88, LB90], DCT [Wat93, HK02, YLL03, YLL05, ZLX05, ZLX08], wavelet transform [TH94a, WHM97, LK00], and polar separable wavelet transform [Wat87, TH94b]. To combine the error of each spatial frequency, orientation band, and location into a single number or a distortion map [Wat79, RG81], many image quality metrics and PICs implement error pooling after CSF, luminance adaptation, and contrast masking.

This chapter begins with a review of the concepts on psychophysics of the human vision that are applied to image quality metrics and computational models for JND. In particular, this chapter emphasizes on image quality metrics and

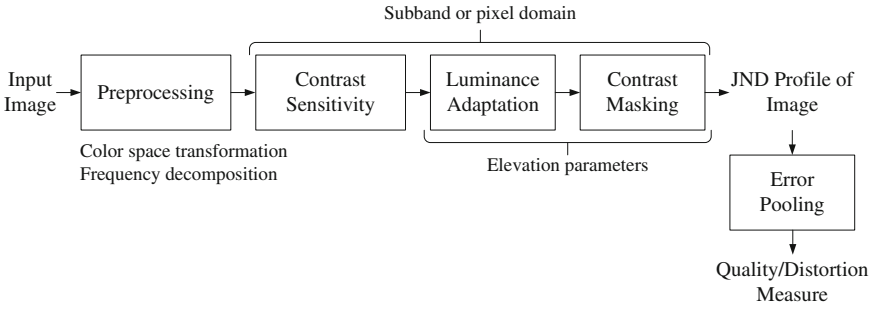

Fig. 2.1 Block diagram of a computational model for JND

computational models for JND that are designed specifically for image compression. As DCT is widely used in image and video compression standards (e.g. JPEG, MPEG-1/2/4, H.261/3), we focus our discussion in this chapter on image quality metrics computed using DCT subbands. It is also useful to consider pixel-based image quality metrics since it is possible to convert the contrast sensitivity from the pixel domain to the DCT subband domain and vice versa.

The cortex filters, which provide a good approximation of the multi-channel response of HVS, and the frequency decomposition using cortex filters are introduced in Sect. 2.1. This is followed by mapping of the cortex filters to DCT-II subbands (or coefficients). In Sect. 2.2, the widely adopted spatial CSF proposed by Ahumada and Peterson [AP92] is discussed. The detection threshold of every DCT subband is inversely proportional to contrast sensitivity, and can be derived from the spatial CSF. Apart from the contrast sensitivity, the detection threshold is varied by the local mean luminance and local spatial content, which are referred as luminance adaptation and contrast masking, respectively. Section 2.3 illustrates the effects of luminance adaptation using Weber's law. Subsequently, several techniques estimating luminance adaptation from the subbands and pixels of an image are reviewed. Next, intra- and inter-band contrast masking are discussed in Sect. 2.4. Intra-band contrast masking is typically adopted in many PICs due to its simple formulation. Discussions on estimating inter-band masking using cortex filters and block classification are also included. The final step of many image quality metrics, known as error pooling, is presented in Sect. 2.5.

2.1 Frequency Decomposition

The multi-channel response of HVS approximates a dyadic system [Dau80] that is well-matched by a multi-resolution filterbank or a wavelet decomposition. Examples of multi-resolution filterbank are cortex transform [Wat87] and cortex filter [Dal92, Dal93]. The cortex transform was first conceived by Watson [Wat87], which was inspired by neurophysiology [HW62, DAT82] and psychophysical studies in masking [BC69, SJ72]. The cortex transform is then adapted by Daly as the cortex filters in VDP. The decomposition of the frequency plane adopted by Watson and Daly is shown in Fig. 2.2. The main difference between Watson's and Daly's implementations of the cortex filtering is that Daly used six orientation bands [PDT77, DYH82], instead of four (in the case of Watson's cortex transform), to better approximate the orientation selectivity of the HVS. Several HVS models [WR84, Wat87, Dal92, Dal93] use six spatial channels, and it was found that six spatial channels show good agreement with psychophysical data [WLM90].

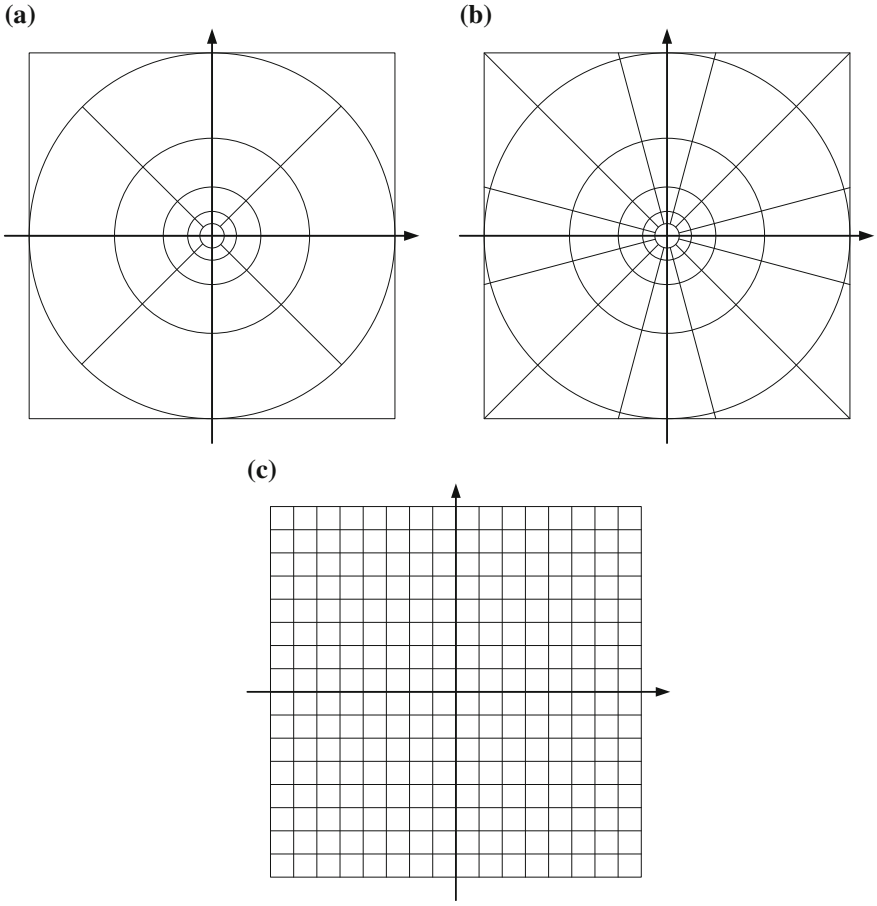

Fig. 2.2 Decomposition of the frequency plane corresponding to **a** Watson's cortex transform [Wat87]. **b** Daly's cortex filters [Dal92, Dal93], and **c** DCT-II. Range of each axis is from $-f_s/2$ to $f_s/2$, where f_s is the sampling frequency

2.1.1 Cortex Filters

The cortex filters model the spatial (or radial) frequency selectivity and the orientational selectivity of the HVS. These filters are formed by cascading two filters, which model the radial frequency bands and orientation bands of the HVS. The radial frequency filters are formed by the difference of two dimensional (2-D) low-pass mesa filters. The mesa filter possesses a flat passband, a Hanning window transition band, and a flat stopband as shown in Fig. 2.3.

The mesa filter [Dal92] is completely characterized by its half-amplitude frequency $d_{1/2}$ and transition width tw. Let s denote the spatial frequency in cycles per degree (cpd). The mesa filter mesa(s) is expressed as

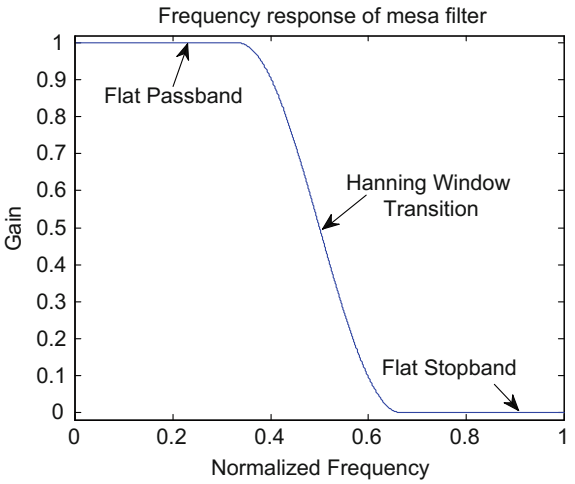

Fig. 2.3 Frequency response of a mesa filter [Dal92, Dal93]

$$\text{mesa}(s) = \begin{cases} 1, & \text{for } s < s_{1/2} - \frac{tw}{2}, \\ \frac{1}{2}\left(1 + \cos\left(\frac{\pi\left(s - s_{1/2} - tw/2\right)}{tw}\right)\right), & \text{for } s_{1/2} - \frac{tw}{2} \le s \le d_{1/2} + \frac{tw}{2}, \\ 0, & \text{for } s > s_{1/2} + \frac{tw}{2}, \end{cases} \quad (2.1)$$

where $tw = 2s_{1/2}/3$. The radial frequency selectivity of the HVS is modelled by the difference of two mesa filters with different half amplitude frequencies. The difference of the mesa (DOM) filter $\text{dom}(d, s)$ is given by

$$\text{dom}(d, s) = \text{mesa}(s)|_{s_{1/2}=2^{-(d-1)}} - \text{mesa}(s)|_{s_{1/2}=2^{-d}}, \quad (2.2)$$

where $d = 0, 1, \ldots, D - 1$, and D is the number of DOM filters. The choice of tw yields a set of cortex bands with approximately constant behaviour on a log frequency axis with a bandwidth of one octave [SJ72, MTT78, DAT82]. The orientation sensitivity of the HVS can be modelled by a set of fan filters [Dal92], which is expressed as

$$\text{fan}(f, \theta) = \begin{cases} \frac{1}{2}\left(1 - \cos\left(\frac{\pi|\theta - \theta_{cr}(f)|}{\theta_{tw}}\right)\right), & |\theta - \theta_c(f)| \le \theta_{tw}, \\ 0, & \text{otherwise}, \end{cases} \quad (2.3)$$

where θ_{tw} is the angular transition width in degree; $\theta_{cr}(f)$ is the orientation of the center angular frequency of the fth fan filter in degree, $f = 0, 1, \ldots, F - 1$, and F is the number of fan filters. $\theta_{cr}(f)$ is given by

$$\theta_{cr}(f) = (f - 1)\theta_{tw} - 90^\circ, \quad (2.4)$$

where $\theta_{tw} = 180°/F$. The cortex filter at the **b**th band cortex$(\mathbf{b}, s, \theta)$ is formed by the product of the dth DOM and fth fan filters, which is given as

$$\text{cortex}(\mathbf{b}, s, \theta) = \begin{cases} \text{dom}(d, s)\text{fan}(f, \theta), & \text{for } d = 1, \ldots, D-1;\ f = 0, 1, \ldots F-1, \\ \text{base}(s), & \text{for } d = D, \end{cases} \tag{2.5}$$

where $\mathbf{b} = (d, f)$, and base(s) is the cortex filter having the lowest spatial frequency without orientational selectivity. In [TS96], the base(s) filter is implemented using a truncated Gaussian function, which is given as

$$\text{base}(s) = \begin{cases} e^{-\frac{s^2}{2\sigma^2}}, & \text{for } s < s_{1/2} + \frac{tw}{2}, \\ 0, & \text{for } s \geq s_{1/2} + \frac{tw}{2}, \end{cases} \tag{2.6}$$

where $\sigma = (2^{-D} + tw/2)/3$. Six spatial channels ($D$ = 6) and six orientation bands (F = 6) are used in Daly's implementation of the cortex filters.

Since the cortex filters model the spatial frequency selectivity and orientation selectivity of the HVS, it would be useful to consider the mapping of DCT-II coefficients to the cortex bands. The general idea behind mapping of DCT-II coefficients to the cortex bands is to group the DCT-II coefficients that belong to the same cortex bands [TS96].

An example of this mapping is illustrated in Fig. 2.4. In order to map the partially covered DCT-II coefficients that fall within a cortex band, Tran and Safranek divide each DCT-II coefficients into $M \times M$ smaller blocks (referred as sub-bins). Subsequently, these sub-bins are grouped into corresponding **b**th band of the cortex filters. Let **k** denote the **k**th DCT-II coefficient of an $N \times N$ DCT-II block, where $\mathbf{k} = (k_1, k_2)$ and $k_1, k_2 = 0, 1, \ldots, N-1$. The overlapping area between the **k**th DCT coefficient and the corresponding band cortex band is computed as

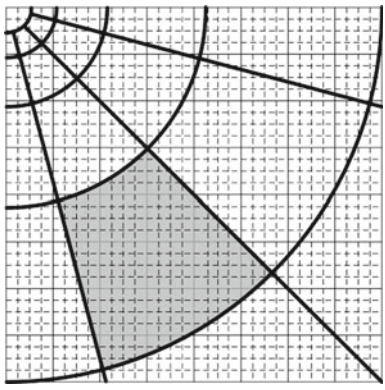

Fig. 2.4 Mapping of DCT coefficients (*thin line*) to cortex bands (*thick line*) [TS96]. The *shaded area* denotes the DCT coefficients which fall within the same cortex band. *Dashed lines* denote the sub-bins of each DCT coefficient

$$\text{overlap}(\mathbf{k},\mathbf{b}) = \sum_{m_1=k_1M}^{(k_1+1)M} \sum_{m_2=k_2M}^{(k_2+1)M} \text{cortex}(\mathbf{b}, m_1, m_2), \tag{2.7}$$

which leads to T_{CF} $N \times N$ matrices, where T_{CF} is the number of cortex filters. Each T_{CF} matrix contains the information of the overlapping area of the N^2 DCT-II coefficients.

2.2 Spatial Contrast Sensitivity Function

In this section, we shall review the widely adopted spatial CSF [Wat93, HK02, YLL03, YLL05, ZLX05, ZLX08], which was proposed by Ahumada and Peterson [AP92]. Their formulation of the CSF is very useful as it takes into account of display luminance levels, veiling luminance levels, and spatial frequencies.

We consider the base detection threshold $T_{\mathrm{D}}(\mathbf{k},\mathbf{n})$ of the $\mathbf{k}$th DCT-II subband located at $\mathbf{n}$ of an image, where $\mathbf{n} = (n_1, n_2)$; $n_1 = 0, 1, \ldots, H/N-1$; $n_2 = 0, 1, \ldots, W/N-1$; H denotes the height of an image and W denotes the width of an image. Let $f(\mathbf{k})$ and $\theta(\mathbf{k})$ denote the spatial frequency of a grating and the angle between two gratings, respectively. Based on van Nes and Bouman's measurements [NB67], Ahumada and Peterson approximated the detection threshold using a parabola in log spatial frequency, and they expressed the detection threshold $T_D(\mathbf{k},\mathbf{n})$ as

$$\log_{10}(T_D(\mathbf{k},\mathbf{n})) = \log_{10}\left(\frac{T_{\min}(\mathbf{n})}{0.7 + 0.3\cos^2\theta(\mathbf{k})}\right) + K(\mathbf{n})(\log_{10} f(\mathbf{k}) - \log_{10} f_{\min}(\mathbf{n}))^2, \quad k_1 = 0 \text{ or } k_2 = 0, \tag{2.8}$$

where $\theta(\mathbf{k}) = \sin^{-1}\left(2f(k_1,0)f(0,k_2)/f^2(\mathbf{k})\right)$, $f(\mathbf{k}) = \sqrt{(k_1/w_x)^2 + \left(k_2/w_y\right)^2}$, w_x and w_y are the horizontal width and vertical height of a pixel, respectively. $T_{\min}(\mathbf{n})$, $K(\mathbf{n})$, and $f_{\min}(\mathbf{n})$ are the functions of the total luminance $L(\mathbf{n})$, where $L(\mathbf{n})$ is the sum of veiling luminance and the luminance of the image located at $\mathbf{n}$. Based on Ahumada and Peterson's formulation, $T_{\min}(\mathbf{n})$, $K(\mathbf{n})$, and $f_{\min}(\mathbf{n})$ are computed as

$$T_{\min}(\mathbf{n}) = \begin{cases} 0.0263L(\mathbf{n})^{0.649}, & L(\mathbf{n}) \le 13.45 \text{ cd/m}^2, \\ 0.0106L(\mathbf{n}), & \text{otherwise}, \end{cases} \tag{2.9}$$

$$f_{\min}(\mathbf{n}) = \begin{cases} 2.401L(\mathbf{n})^{0.182}, & L(\mathbf{n}) \le 300 \text{ cd/m}^2, \\ 6.78, & \text{otherwise}, \end{cases} \tag{2.10}$$

and

$$K(\mathbf{n}) = \begin{cases} 2.0891 L(\mathbf{n})^{0.0706}, & L(\mathbf{n}) \le 300 \text{ cd/m}^2, \\ 3.125, & \text{otherwise.} \end{cases} \tag{2.11}$$

Since van Nes and Bouman found negligible difference between the CSFs for luminance ranging from 290 to 1880 cd/m^2, (2.10) and (2.11) are clipped at 300 cd/m^2. It should be noted that Kelly [Kel85] stated that this parabola model of the CSF may not be valid for low spatial frequencies, and Peterson et al. [PMP91] suggested a conservative estimate for $T_D(0,0,\mathbf{n})$, which is the smaller value of $T_D(1,0,\mathbf{n})$ and $T_D(0,1,\mathbf{n})$.

Watson [Wat93] used the DC DCT-II coefficient to estimate the local luminance of an image. Höntsch and Karam [HK02] estimated the local luminance from the foveal region, which typically covers two degrees of the visual angle, as

$$L(\mathbf{n}) = L_{\min} + \frac{L_{\max} - L_{\min}}{M} \left(\sum_{(0,0,m_1,m_2) \in F(0,0,\mathbf{n})} \frac{C(0,0,m_1,m_2)}{N_F N} + \bar{m} \right), \tag{2.12}$$

where $F(0,0,\mathbf{n})$ denotes the foveal region centers at $\mathbf{n}$ in DC subband; $C(0,0,m_1,m_2)$ denotes the DC DCT-II coefficient at (m_1,m_2); N_F denotes the number of DCT-II coefficients at $\mathbf{n}$ in DC subband that fall inside the foveal region; and $\bar{m}$ is the mean of the image; M is the number of gray levels in the image; $L_{\max}$ and $L_{\min}$ are the maximum and minimum luminance levels of the display, respectively. N_F is computed as

$$N_F = \left(\left\lfloor \frac{2VR_x}{N} \tan\left(\frac{\theta_f}{2}\right) \right\rfloor \right) \left(\left\lfloor \frac{2VR_y}{N} \tan\left(\frac{\theta_f}{2}\right) \right\rfloor \right), \tag{2.13}$$

where the operator $\lfloor . \rfloor$ returns the nearest smallest integer; V is the viewing distance in inches; R_x and R_y are the height and width of the display resolution in pixel per inch, respectively; and θ_f is the visual angle (approximately 2°) covered by the foveal region.

Assuming an image is displayed on a gamma corrected screen, we can linearly map signal intensity values into luminance levels. Thereby, the base detection threshold $T_{\mathrm{b}}(\mathbf{k},\mathbf{n})$ for the $\mathbf{k}$th DCT-II subband located at $\mathbf{n}$ is computed as

$$T_{\mathrm{b}}(\mathbf{k},\mathbf{n}) = \frac{M T_{\mathrm{D}}(\mathbf{k},\mathbf{n})}{\alpha_{k_1}\alpha_{k_2}(L_{\max} - L_{\min})}. \tag{2.14}$$

To ensure the quantization error remains invisible to the HVS, the quantization of each DCT-II coefficient should not be greater than $2T_{\mathrm{b}}(\mathbf{k},\mathbf{n})$.

The JND threshold for DCT-II subband is formulated as a product of the detection threshold $T_{\mathrm{b}}(\mathbf{k},\mathbf{n})$ and its elevation parameters given by luminance adaptation and contrast masking. Let $e_{\mathrm{la}}(\mathbf{n})$ and $e_{\mathrm{cm}}(\mathbf{k},\mathbf{n})$ denote the luminance

adaption and contrast masking, respectively. Hence, the JND threshold $T(\mathbf{k}, \mathbf{n})$ for the $\mathbf{k}$th DCT-II subband located at $\mathbf{n}$ is given as

$$T(\mathbf{k}, \mathbf{n}) = T_{\mathrm{b}}(\mathbf{k}, \mathbf{n}) e_{\mathrm{la}}(\mathbf{n}) e_{\mathrm{cm}}(\mathbf{k}, \mathbf{n}). \tag{2.15}$$

Since the luminance of a digital image spans a small luminance range of the spatial CSF experiment conducted by van Nes and Bouman [NB67], a single spatial CSF (based on the mean luminance of the image) can be used for the whole image [ZLX05]. Therefore, the detection threshold $T_{\mathrm{D}}(\mathbf{k}, \mathbf{n})$ can be simplified to $T_{\mathrm{D}}(\mathbf{k})$ by replacing the total luminance $L(\mathbf{n})$ with the mean luminance L of the display [Wat93, ZLX05, ZLX08].

2.3 Luminance Adaptation

Weber's law is widely used to model luminance adaptation, and the Weber fraction $K = \Delta I / I_{bg}$ is found to be nearly constant for a wide range of intensities [Hec24], where I_{bg} is the background intensity and ΔI is the just-noticeable incremental intensity over the background. However, Weber's law does not hold for a wide range of background intensities and spatial frequencies. For an 8-bit grayscale image, it is found that the Weber's fraction stays fairly constant for gray levels from 50 to 235; and higher contrast sensitivity [SW96] is found for gray levels lower and higher than 50 and 235, respectively. These observations are similar to those reported in [SJ89, CL95]. From the empirical model of the CSF in [Bar04], it is also observed that the contrast sensitivity remains relatively constant at low spatial frequencies for luminance levels between 10 and 1000 cd/m^2. However, the contrast sensitivity for these luminance levels vary significantly as the spatial frequency increases.

In the DCT domain, Watson [Wat93] estimated the luminance adaptation for $\mathbf{n}$th DCT-II block using

$$e_{\mathrm{la}}^{\mathrm{Wat}}(\mathbf{n}) = \left(\frac{C(0,0,\mathbf{n})}{\bar{C}_L} \right)^{0.649}, \tag{2.16}$$

where $\bar{C}_L$ refers to the DC DCT-II coefficient corresponding to the mean luminance ($\bar{C}_L$ = 1024 for a 8-bit image). On the other hand, Zhang et al. [ZLX05, ZLX08] considered different luminance adaptation at low and high luminance, and they estimated the luminance adaptation as

$$e_{\mathrm{la}}^{\mathrm{ZLX}}(\mathbf{n}) = \begin{cases} 2\left(1 - \frac{C(0,0,\mathbf{n})}{128N}\right)^3 + 1, & \text{for } C(0,0,\mathbf{n}) \le 128N, \\ 0.8\left(\frac{C(0,0,\mathbf{n})}{128N} - 1\right)^2 + 1, & \text{otherwise.} \end{cases} \tag{2.17}$$

Using a similar approach, Wei and Ngan [WN09] computed luminance adaptation using

$$e_{\mathrm{la}}^{\mathrm{WN}}(\mathbf{n}) = \begin{cases} \left(\frac{60N - C(0,0,\mathbf{n})}{150N}\right) + 1, & \text{for } C(0,0,\mathbf{n}) \leq 60N, \\ 1, & \text{for } 60N < C(0,0,\mathbf{n}) < 170N, \\ \left(\frac{C(0,0,\mathbf{n}) - 170N}{425N}\right) + 1, & \text{for } C(0,0,\mathbf{n}) \geq 170N. \end{cases} \tag{2.18}$$

In the pixel domain, Chou and Li [CL95, YLL05] empirically determined the luminance adaptation of a pixel at $\mathbf{x}$, where $\mathbf{x} = (x_1, x_2)$, $x_1 = 0, 1, \ldots, H-1$, and $x_2 = 0, 1, \ldots, W-1$, using the following:

$$e_{\mathrm{la}}^{\mathrm{CL}}(\mathbf{x}) = \begin{cases} 17\left(1 - \sqrt{\frac{L_s(\mathbf{x})}{127}}\right) + 3, & \text{for } L_s(\mathbf{x}) \leq 127, \\ \frac{3}{128}\left(L_s(\mathbf{x}) - 127\right) + 3, & \text{for } L_s(\mathbf{x}) > 127, \end{cases} \tag{2.19}$$

where $L_s(\mathbf{x})$ is the local luminance at $\mathbf{x}$, and (2.19) was obtained for a distance of six times of the image height. Chou and Li determined the local luminance $L_s(\mathbf{x})$ as

$$L_s(\mathbf{x}) = \frac{1}{32} \sum_{p_1=0}^{4} \sum_{p_2=0}^{4} i(x_1 - 2 + p_1, x_2 - 2 + p_2) B(p_1, p_2), \tag{2.20}$$

where $i(\mathbf{x})$ denotes the pixel of an image at $\mathbf{x}$ and the operator B is depicted in Fig. 2.5.

Fig. 2.5 Operator to determine average local luminance (B) [CL95]

1	1	1	1	1
1	2	2	2	1
1	2	0	2	1
1	2	2	2	1
1	1	1	1	1

B

2.4 Contrast Masking

Contrast masking refers to the reduction of visibility of one image signal due to the presence of another signal. The masking characteristic of the HVS is known to be strongest when both signals are of the same spatial frequency, orientation, and location [LF80]. Contrast masking can be classified as inter- and intra-band masking. Sometimes, the term "texture masking" (inter-band masking) is used to refer to a "broadband" masker, where the masking effect is contributed by multiple frequency and orientation channels. On the other hand, intra-band masking refers to the masking due to a masker within the same frequency and orientation channel. Based on the estimation of contrast masking reported in [SJ89], Höntsch and Karam [HK00] proposed a more elaborate adjustment for contrast masking, which incorporates both intra- and inter-band masking. Let $e_{\text{inter}}(\mathbf{k},\mathbf{n})$ and $e_{\text{intra}}(\mathbf{k},\mathbf{n})$ denote the amount of intra- and inter-band masking at $\mathbf{n}$ of $\mathbf{k}$th subband of an image, respectively. The elevation parameter $e_{\text{cm}}(\mathbf{k},\mathbf{n})$ is computed as

$$e_{\text{cm}}(\mathbf{k},\mathbf{n}) = e_{\text{inter}}(\mathbf{k},\mathbf{n}) e_{\text{intra}}(\mathbf{k},\mathbf{n}). \tag{2.21}$$

In [SJ89], Safranek and Johnston proposed a subband image coder that employs a 4×4 band generalized QMF (GQMF) to decompose an image into 16 subbands. Let $tex^{\text{SJ}}(\mathbf{b},\mathbf{n}')$ denote the texture energy of the $\mathbf{b}$th subband at location $\mathbf{n}'$, and $\text{wCSF}(\mathbf{b})$ is the $\mathbf{b}$th weighting factor empirically derived from a CSF [Cor90], where $\mathbf{n}' = (n_1', n_2')$, $n_1' = 0, 1, \ldots, H/2-1$, $n_2' = 0, 1, \ldots, W/2-1$, $\mathbf{b} = (b_1, b_2)$, and $0 \le b_1, b_2 \le 3$. Safranek and Johnston defined contrast masking (only inter-band masking is considered) as follows:

$$e_{\text{inter}}^{\text{SJ}}(\mathbf{b},\mathbf{n}') = \max\left\{1, \left(\sum_{\mathbf{b}} \text{wCSF}(\mathbf{b}) tex^{\text{SJ}}(\mathbf{b},\mathbf{n}')\right)^{0.15}\right\}. \tag{2.22}$$

The texture energy of the $\mathbf{b}$th subband at location $\mathbf{n}'$ is computed as

$$tex^{\text{SJ}}(\mathbf{b},\mathbf{n}') = \begin{cases} \text{var}(\mathbf{n}'), & \text{for } \mathbf{b} = (0,0), \\ \text{energy}(\mathbf{b},\mathbf{n}'), & \text{otherwise}, \end{cases} \tag{2.23}$$

where $\text{energy}(\mathbf{b},\mathbf{n}')$ computes the energy of the $\mathbf{b}$th subband at $\mathbf{n}'$ and $\text{var}(\mathbf{n}')$ computes the variance at $\mathbf{n}'$ of subband zero over the area given by (n_1', n_2'), $(n_1'+1, n_2')$, $(n_1', n_2'+1)$, and $(n_1'+1, n_2'+1)$.

Based on the masking model in [LF80], Watson [Wat93] adjusted the base detection threshold to account for contrast masking (only intra-band masking is considered) using the following:

$$e_{\text{intra}}^{\text{Wat}}(\mathbf{k},\mathbf{n}) = \begin{cases} 1, & \mathbf{k} = (0,0), \\ \max\left\{1, \left(\frac{|C(\mathbf{k},\mathbf{n})|}{T_{\text{b}}(\mathbf{k},\mathbf{n}) e_1^{\text{Wat}}(\mathbf{n})}\right)^{0.7}\right\}, & \mathbf{k} \neq (0,0). \end{cases} \tag{2.24}$$

It is assumed that there is no contrast masking in the DC DCT-II coefficient. However, the DC DCT-II coefficient indirectly affects contrast masking via $e_1^{\text{Wat}}(\mathbf{n})$ in the denominator of (2.24) for all DCT-II coefficients except for the DC DCT-II coefficient.

In [HK00], Höntsch and Karam estimated intra-band masking using

$$e_{\text{intra}}^{\text{HK}}(\mathbf{k},\mathbf{n}) = \begin{cases} 1, & \mathbf{k} = (0,0), \\ \max\left\{1, \left(\frac{|C(\mathbf{k},\mathbf{n})|}{T_{\text{b}}(\mathbf{k},\mathbf{n})}\right)^{0.36}\right\}, & \mathbf{k} \neq (0,0), \end{cases} \tag{2.25}$$

and inter-band masking is computed using (2.22) with an exponent of 0.035. Taking into account of the foveal region for intra-band masking, Höntsch and Karam [HK02] proposed the adjustment for intra-band masking as

$$e_{\text{intra}}^{\text{HK2}}(\mathbf{k},\mathbf{n}) = \begin{cases} 1, & \mathbf{k} = (0,0), \\ \max\left\{1, \left(\frac{|\bar{C}_F(\mathbf{k},\mathbf{n})|}{T_{\text{b}}(\mathbf{k},\mathbf{n})}\right)^{0.6}\right\}, & \mathbf{k} \neq (0,0), \end{cases} \tag{2.26}$$

where $\bar{C}_F(\mathbf{k},\mathbf{n})$ is the average magnitude of the DCT-II coefficients in the foveal region.

Yang et al. [YLL05] improved their estimate of contrast masking by differentiating the contribution of masking from edge and texture. Edges are structurally simpler than textures, and it is generally observed that edges tend to be easily recognized by the HVS. Furthermore, a typical observer would have prior knowledge of how an edge looks like [EB98]. Girod [Gir93] found that the HVS has acute sensitivity at or near the luminance edge. Based on these observations in [EB98, Gir93], Yang et al. defined the JND threshold at a texture region to be three times higher than those at an edge region.

To date, classification of plain, edge, and texture blocks are performed in [YLL05, ZLX05, ZLX08, WN09, LLP10] to effectively estimate contrast masking in an image. Zhang et al. [ZLX05, ZLX08] employed a block classification method in the DCT domain [TV98], which was first proposed in [PJJ94]. To perform block classification in the DCT domain, the DCT-II coefficients of an $N \times N$ sub-image are divided into four groups as shown in Fig. 2.6. Let $L_{\text{T}}(\mathbf{n})$, $M_{\text{T}}(\mathbf{n})$, and $H_{\text{T}}(\mathbf{n})$ denote the sum of DCT-II coefficients (absolute magnitude) in the low-frequency (LF), mid-frequency (MF), and high-frequency (HF) groups, respectively, of the **n**th DCT-II block. Based on these sums, three measures are formulated to determine the texture energy of the **n**th DCT-II block, and these measures are defined as

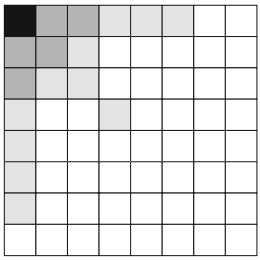

Fig. 2.6 DCT-II block classification for contrast masking. LF, MF, and HF are represented by the *dark gray*, *light gray* and *white boxes*, respectively [ZLX05, ZLX08]

$$
\begin{aligned}
tex_1^{\mathrm{ZLX}}(\mathbf{n}) &= M_{\mathrm{T}}(\mathbf{n}) + H_{\mathrm{T}}(\mathbf{n}),\\
tex_2^{\mathrm{ZLX}}(\mathbf{n}) &= \frac{(\bar{L}_{\mathrm{T}}(\mathbf{n}) + \bar{M}_{\mathrm{T}}(\mathbf{n}))}{\bar{H}_{\mathrm{T}}(\mathbf{n})},\\
tex_3^{\mathrm{ZLX}}(\mathbf{n}) &= \frac{\bar{L}_{\mathrm{T}}(\mathbf{n})}{\bar{M}_{\mathrm{T}}(\mathbf{n})},
\end{aligned}
\tag{2.27}
$$

where $\bar{L}_{\mathrm{T}}(\mathbf{n})$, $\bar{M}_{\mathrm{T}}(\mathbf{n})$, and $\bar{H}_{\mathrm{T}}(\mathbf{n})$ are the means of $L_{\mathrm{T}}(\mathbf{n})$, $M_{\mathrm{T}}(\mathbf{n})$, and $H_{\mathrm{T}}(\mathbf{n})$, respectively.

Each DCT-II block is classified into PLAIN, EDGE, or TEXTURE class using $tex_1^{\mathrm{ZLX}}(\mathbf{n})$, $tex_2^{\mathrm{ZLX}}(\mathbf{n})$, and $tex_3^{\mathrm{ZLX}}(\mathbf{n})$ as shown in Table 2.1. DCT-II blocks that are generally smooth with few spatial activities are classified as PLAIN, DCT-II blocks containing a lot of complex spatial activities are classified as TEXTURE, and DCT-II blocks containing clear edges are classified as EDGE.

Based on the block classification result, inter-band contrast masking is computed by

Table 2.1 Conditions used in classification of DCT-II blocks [ZLX05, ZLX08]

Case	Conditions	Block classification
I	$tex_1^{\mathrm{ZLX}}(\mathbf{n}) \le 125$	DCT-II block is classified as PLAIN
II	$125 < tex_1^{\mathrm{ZLX}}(\mathbf{n}) \le 290$ and $\max\left(tex_2^{\mathrm{ZLX}}(\mathbf{n}), tex_3^{\mathrm{ZLX}}(\mathbf{n})\right) \ge 7$ $\min\left(tex_2^{\mathrm{ZLX}}(\mathbf{n}), tex_3^{\mathrm{ZLX}}(\mathbf{n})\right) \ge 5$ or $tex_2^{\mathrm{ZLX}}(\mathbf{n}) \ge 16$	DCT-II block is classified as EDGE, otherwise PLAIN
III	$290 < tex_1^{\mathrm{ZLX}}(\mathbf{n}) \le 900$ and $\max\left(tex_2^{\mathrm{ZLX}}(\mathbf{n}), tex_3^{\mathrm{ZLX}}(\mathbf{n})\right) \ge 7$ $\min\left(tex_2^{\mathrm{ZLX}}(\mathbf{n}), tex_3^{\mathrm{ZLX}}(\mathbf{n})\right) \ge 5$ or $tex_2^{\mathrm{ZLX}}(\mathbf{n}) \ge 16$	DCT-II block is classified as EDGE, otherwise TEXTURE
IV	$tex_1^{\mathrm{ZLX}}(\mathbf{n}) > 900$ and $\max\left(tex_2^{\mathrm{ZLX}}(\mathbf{n}), tex_3^{\mathrm{ZLX}}(\mathbf{n})\right) \ge 0.7$ $\min\left(tex_2^{\mathrm{ZLX}}(\mathbf{n}), tex_3^{\mathrm{ZLX}}(\mathbf{n})\right) \ge 0.5$ or $tex_2^{\mathrm{ZLX}}(\mathbf{n}) \ge 16$	DCT-II block is classified as EDGE, otherwise TEXTURE

$$e_{\text{inter}}^{\text{ZLX}}(\mathbf{n}) = \begin{cases} 1 + \frac{tex_{\text{l}}^{\text{ZLX}}(\mathbf{n}) - 290}{1208}, & \text{for TEXTURE block,} \\ 1.25, & \text{for EDGE block and } L(\mathbf{n}) + M(\mathbf{n}) > 400, \\ 1.125, & \text{for EDGE block and } L(\mathbf{n}) + M(\mathbf{n}) \le 400, \\ 1, & \text{for PLAIN block.} \end{cases} \tag{2.28}$$

Zhang et al. considered similar adjustment as (2.24) for intra-band contrast masking, and the amount of adjustment for contrast masking is computed as

$$e_{\text{intra}}^{\text{ZLX}}(\mathbf{k}, \mathbf{n}) = \begin{cases} 1, & \text{for EDGE block at } \mathbf{k} \in LF \cup MF, \\ \max\left\{1, \left(\frac{|C(\mathbf{k},\mathbf{n})|}{T_b(\mathbf{k},\mathbf{n}) e_{\text{la}}^{\text{ZLX}}(\mathbf{k},\mathbf{n})}\right)^{0.36}\right\}, & \text{otherwise.} \end{cases} \tag{2.29}$$

To avoid over-estimation of JND threshold at the EDGE block, the LF and MF regions of the EDGE block are excluded from the estimation of intra-band contrast masking.

Differing from Zhang's method, Wei and Ngan [WN09] performed block classification in the pixel domain. Using an edge map of the image obtained with the Canny edge detector [Can86], Wei and Ngan computed the edge density $\bar{p}_{\text{edge}}(\mathbf{n})$ at $\mathbf{n}$ as the ratio of the number of edge pixels in each $N \times N$ sub-image to N^2. Based on the edge density, the $\mathbf{n}$th DCT-II block is classified as

$$\text{Block Type}(\mathbf{n}) = \begin{cases} \text{PLAIN} & \text{for } \bar{p}_{\text{edge}}(\mathbf{n}) \le 0.1, \\ \text{EDGE} & \text{for } 0.1 < \bar{p}_{edge}(\mathbf{n}) \le 0.2, \\ \text{TEXTURE} & \text{for } \bar{p}_{\text{edge}}(\mathbf{n}) > 0.2. \end{cases} \tag{2.30}$$

Using the block classification results from (2.30), the inter-band masking is computed as

$$e_{\text{inter}}^{\text{WN}}(\mathbf{k}, \mathbf{n}) = \begin{cases} 1 & \text{for PLAIN and EDGE block,} \\ 2.25 & \text{for } \left(k_1^2 + k_2^2\right) \le 16 \text{ in TEXTURE block,} \\ 1.25 & \text{for } \left(k_1^2 + k_2^2\right) > 16 \text{ in TEXTURE block.} \end{cases} \tag{2.31}$$

Finally, Wei and Ngan computed intra-band contrast masking as

$$e_{\text{intra}}^{\text{WN}}(\mathbf{k}, \mathbf{n}) = \begin{cases} 1, & \text{for } \left(k_1^2 + k_2^2\right) \le 16 \text{ in PLAIN and EDGE block,} \\ \min\left\{4, \max\left\{1, \left(\frac{|C(\mathbf{k},\mathbf{n})|}{T_b(\mathbf{k},\mathbf{n}) e_{\text{la}}^{\text{WN}}(\mathbf{k},\mathbf{n})}\right)^{0.36}\right\}\right\}, & \text{otherwise.} \end{cases} \tag{2.32}$$

0	0	0	0	0
1	3	8	3	1
0	0	0	0	0
-1	-3	-8	-3	1
0	0	0	0	0

G_1

0	0	1	0	0
0	8	3	0	0
1	3	0	-3	1
0	0	-3	-8	0
0	0	-1	0	0

G_2

0	0	1	0	0
0	0	3	8	0
1	3	0	-3	1
0	-8	-3	0	0
0	0	-1	0	0

G_3

0	1	0	-1	0
0	3	0	-3	0
1	8	0	-8	1
0	3	0	-3	0
0	1	0	-1	0

G_4

Fig. 2.7 Operators to determine weighted average of luminance changes (G_p)

For a viewing distance of six times of the image height, Chou and Li [CL95] estimated contrast masking in the pixel domain using the following expression:

$$e_2^{\mathrm{CL}}(\mathbf{x}) = 0.01L(\mathbf{x})(0.01G(\mathbf{x}) - 1) + 0.115G(\mathbf{x}) + 0.5, \tag{2.33}$$

where $G(\mathbf{x})$ is the maximal weighted average of the gradient around the pixel at $\mathbf{x}$. $G(\mathbf{x})$ is calculated by

$$G(\mathbf{x}) = \max_{j=1,2,3,4} \left\{ \left| grad_j(\mathbf{x}) \right| \right\}, \tag{2.34}$$

where

$$grad_j(\mathbf{x}) = \frac{1}{16} \sum_{p_1=0}^{4} \sum_{p_2=0}^{4} c(x_1 - 2 + p_1, x_2 - 2 + p_2) G_j(p_1, p_2), \tag{2.35}$$

and $G_j(\mathbf{x})$ are the four directional highpass filters shown in Fig. 2.7.

2.5 Error Pooling

The final step of many image quality metrics is to combine the errors normalized by $T(\mathbf{k}, \mathbf{n})$ computed for every spatial frequency (from DCT-II subbands) at all spatial location $\mathbf{n}$ into a single distortion measure [SJ89, Wat93]. Alternatively, these normalized errors can be combined into an error map using error pooling, which describes the amount of error of each pixel in the image.

An example of error pooling using the Minkowski metric can be expressed as

$$P(\mathbf{n}) = \left(\sum_{\mathbf{k}} \left| \frac{C(\mathbf{k}, \mathbf{n}) - \hat{C}(\mathbf{k}, \mathbf{n})}{T(\mathbf{k}, \mathbf{n})} \right|^{\beta_f} \right)^{1/\beta_f}, \tag{2.36}$$

where $\hat{C}(\mathbf{k}, \mathbf{n})$ is the quantized $\mathbf{k}$th DCT-II coefficient of the $\mathbf{n}$th DCT-II block and β_f is a constant for summation across frequency bands. For summation across frequency band, it is found that $\beta_f \approx 4$ [Wat82, GRN78, Leg78a, Leg78b, RG81, PAW93b, RAW97]. By summing all the errors in (2.36) over $\mathbf{n}$, a single value describing the distortion of an image is then obtained. For spatial error pooling over $\mathbf{n}$, several values of β_s have been adopted. Teo and Heeger [TH94b], Lubin [Lub93, Lub95], and Watson [Wat93] adopted β_s as 2, 2.4, and 4, respectively. Alternatively, error pooling can be performed over $\mathbf{n}$, followed by over frequency bands [Wat93].

At near JND threshold, probability summation is well accepted as the basis for summation of signal energy (or distortion) across frequency and spatial domains [EB98]. For summation across frequency band, it is reported in [GRN78, Leg78a, Leg78b, RG81] that $\beta_f = 3.5$ [Wat82] is consistent with subjective evaluation. It has been found that summation across frequency bands with DCT-II basis functions is well modeled with $\beta_f = 2.4$ [PAW93b]. In target detection experiments [RAW97], it is found that $\beta_s = 4$ provides the closest match to psychophysical results for spatial summing.

To obtain a single distortion value describing the amount of distortion in a compressed image, spatial summing is performed after summation across frequency bands or vice versa. If summation across frequency bands is first performed, the perceptual distortion score P_1 of an image becomes

$$P_1 = \left(\sum_{\mathbf{n}} P(\mathbf{n})^{\beta_s} \right)^{1/\beta_s}. \tag{2.37}$$

Alternatively, localized pooling of an image can be performed. One such example is found in [HK02], where spatial summing is performed within the foveal region $F(\mathbf{k}, \mathbf{n})$. The distortion within the foveal region is given as

$$P_F(\mathbf{k}, \mathbf{n}) = \left(\sum_{(\mathbf{k}', \mathbf{n}') \in F(\mathbf{k}, \mathbf{n})} \left| \frac{C(\mathbf{k}', \mathbf{n}') - \hat{C}(\mathbf{k}', \mathbf{n}')}{T(\mathbf{k}', \mathbf{n}')} \right|^{\beta_F} \right)^{1/\beta_F}, \tag{2.38}$$

where $\beta_F = 4$. Using the foveal distortion $P_F(\mathbf{k}, \mathbf{n})$, the distortion for the $\mathbf{k}$th DCT-II coefficient is computed as

$$P_F(\mathbf{k}) = \max_{\mathbf{n}} \{ P_F(\mathbf{k}, \mathbf{n}) \}, \tag{2.39}$$

and the single distortion measure of the image becomes

$$P_F = \max_{\mathbf{k}} \{ P_F(\mathbf{k}) \}. \tag{2.40}$$

Zhang et al. [ZLX05, ZLX08] suggested the following expression for spatial summing:

$$P(\mathbf{k}) = \begin{cases} \left(\sum_{\mathbf{n}} |d_{\text{JND}}(\mathbf{k}, \mathbf{n})|^{2.3} \right)^{1/2.3}, & \text{for } \mathbf{k} = (0,0), (1,0), (0,1), \\ \left(\sum_{\mathbf{n}} |d_{\text{JND}}(\mathbf{k}, \mathbf{n})|^{4} \right)^{1/4}, & \text{otherwise,} \end{cases} \tag{2.41}$$

where $d_{\text{JND}}(\mathbf{k}, \mathbf{n}) = \big(C(\mathbf{k}, \mathbf{n}) - \hat{C}(\mathbf{k}, \mathbf{n})\big) \big/ T(\mathbf{k}, \mathbf{n})$. In this case, the perceptual distortion score P_2 is computed as:

$$P_2 = \left(\sum_{\mathbf{k}} P(\mathbf{k})^{\beta_f} \right)^{1/\beta_f}. \tag{2.42}$$

2.6 Summary

In this chapter, we reviewed three properties of the HVS, namely, CSF, luminance adaptation, and intra- and inter-band contrast masking. These properties play important roles in the design of image quality metric and computational model for JND. It is known that DCT does not match the channel decomposition mechanism of the HVS. To mitigate the issues arise from the mismatch of frequency decomposition of the HVS and DCT, Tran and Safranek [TS96] introduced a mapping from DCT-II coefficients to the cortex bands. Section 2.1 introduced the cortex filters, and reviewed the mapping of DCT-II coefficients to the cortex bands. Section 2.2 presented a widely adopted CSF proposed by Ahumada and Peterson [AP92], which is used to compute the base detection threshold of DCT subband.

Elevation in the base detection threshold is attributed by the luminance adaptation and contrast masking. These elevation parameters were reviewed in Sects. 2.3 and 2.4, respectively. Luminance adaptation refers to the variation of the base detection threshold due to the local luminance. Two forms of contrast masking, namely, the intra- and inter-band contrast masking were described in Sect. 2.4. Most PICs account for intra-band contrast masking due to its simple formulation; however more accurate representation of the JND threshold should also include inter-band contrast masking. Two estimations of the inter-band contrast masking using block classification and cortex filtering were shown in Sect. 2.4.

In Sect. 2.5, we discussed how a PIC uses a single distortion measure or distortion map to determine the permissible compression of an entire image (using a single distortion measure) or different regions of the image (using a distortion map) at a predefined image quality. The next chapter shall review the integration of these computational models in DCT-based image coders.

Chapter 3
Perceptual Image Coding with Discrete Cosine Transform

Lossless compression typically offers compression ratio at the order of 3:1 [WSS00]. Since limited storage and transmission bandwidth are available, such compression ratio is inadequate for most applications. While lossy compression overcomes this problem, overly compressed JPEG and JPEG 2000 images exhibit various artifacts, such as blocking and blurring, respectively.

At the same level of compression, JPEG 2000 [CSE00] images do not exhibit the same level of degradation as compared to JPEG [Wal92] images. This is attributed to the discrete wavelet transform (DWT) and DCT-II adopted by the JPEG 2000 [ISO00] and JPEG [ISO94] standards, respectively. At high compression ratio, content of JPEG 2000 images can be easily recognized as compared to JPEG images. Furthermore, JPEG 2000 offers PSNR and resolution scalability. The main disadvantage of JPEG 2000 is its higher complexity as compared to other image compression standards [SE00].

To date, DCT is still popular and used in numerous image and video compression standards, such as JPEG, MPEG-1/2/4, and H.261/3, therefore this chapter focuses on the discussion of perceptually-tuned image coder (PIC) that is based on DCT-II. By exploiting the limitation of HVS, perceptually-tuned image compression improves coding efficiency of images without introducing highly visible compression artifact. Specifically, the physiological and psychological mechanisms of the HVS prevent detection of all changes (such as compression artifact) in an image, and the threshold which defines the minimum sensory difference detectable by the HVS is commonly referred as the JND threshold [JJS93].

Over the years, several computational models for JND have been proposed and these models are computed from subbands [SJ89, Wat93, TS96, HK00, HK02, ZLX05, ZLX08, WN09] or pixels [CL95, CC96, CB99, YLL03, YLL05, LLP10, TG11] of an image. In this chapter, we shall discuss how computational models for JND in the pixel and DCT-II domains are integrated into DCT-II based PICs. We focus on DCT-II based PICs since this class of PIC can be made compatible with the popular JPEG standard. Such integration would permit perceptually better images, especially at low bit coding, while remaining compatible with the JPEG standard.

E.-L. Tan and W.-S. Gan, *Perceptual Image Coding with Discrete Cosine Transform*, SpringerBriefs in Signal Processing,
DOI 10.1007/978-981-287-543-3_3

We shall begin this chapter with a brief introduction of the four modes of operations defined in the JPEG standard. Three PICs integrated with JND models in the DCT-II domain [Wat93, ZLX05, WN09] are presented in Sect. 3.2. This is followed by a comparative analysis and discussions of some key modules of these PICs, such as luminance adaptation and block classification. Section 3.3 reviews the JND models in the pixel domain [CL95, YLL05]. A discussion on the NAMM proposed by Yang et al. [YLL03, YLL05] and its relationship with popular JND models in the pixel domain shall be presented. The steps required to integrate a JND model in the pixel domain to a DCT-II based PIC is also discussed. The computation of the quantization matrices for these two classes of PICs is reviewed in Sect. 3.4. Finally, this chapter is summarized in Sect. 3.5.

3.1 Still-Image Compression Standard—JPEG

Since the mid-1980s, the International Telecommunication Union (ITU) and International Organization for Standardization (ISO) have been working together to establish a compression standards for still images. This joint committee is known as the JPEG, and their collaborative effort led to the development of the ISO/IEC international standard 10918-1: digital compression and coding of continuous-tone still images, or the ITU-T Recommendation T.81.

3.1.1 Modes of JPEG Standard

To meet the requirements of various imagery applications, the JPEG standard defines four modes of operations, namely, sequential DCT-based, progressive DCT-based, lossless, and hierarchical. In the sequential DCT-based mode, the image is divided into 8×8 sub-images from left to right and top to bottom. The 8×8 DCT-II is used to transform each sub-image into 8×8 DCT-II coefficients. Subsequently, these DCT-II coefficients are quantized and then entropy encoded.

In the progressive DCT-based mode, the image is encoded in multiple scans. In other words, the quantized DCT coefficients are partially coded in each scan by either spectral selection or successive approximation. An example of spectral selection and successive approximation is shown in Fig. 3.1. In spectral selection (see Fig. 3.1a), the quantized DCT coefficients are divided into several spectral bands and each band is encoded in each scan. In successive approximation (see Fig. 3.1b), the most significant bits of the quantized DCT coefficients are first encoded and followed by encoding of the lesser significant bits in the subsequent scans. The graphical illustration of the differences between the progressive and

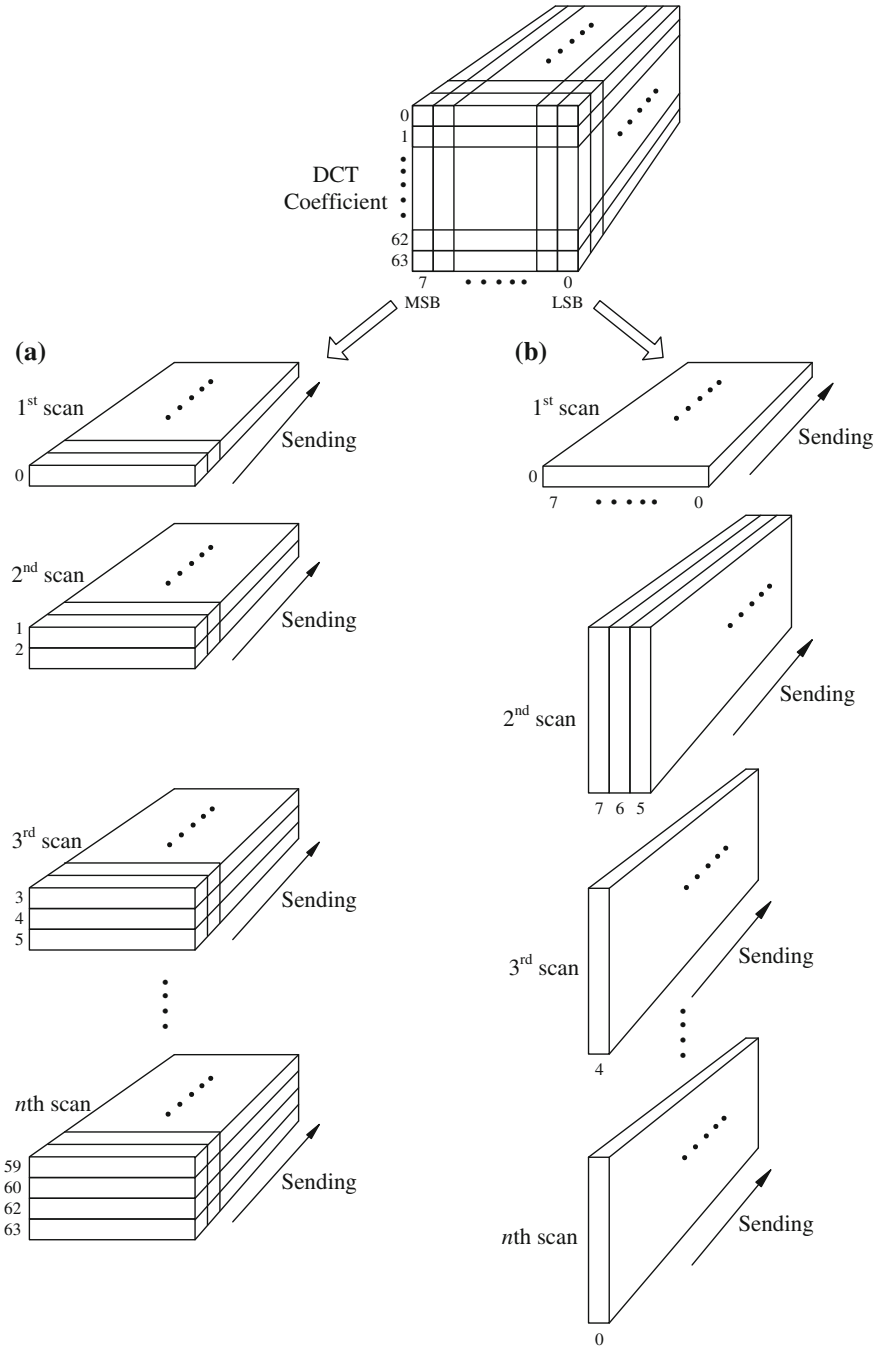

Fig. 3.1 Progressive mode of JPEG. **a** Progressive DCT-based mode with spectrum selection, and **b** successive approximation

sequential DCT-based modes is shown in Fig. 3.2. The progressive mode allows a lower resolution of the image to be viewed without decompressing the image at its full resolution, and permits the image to build up from several coarse-to-clear passes. On the other hand, the image is decoded from the top left to right and top to bottom for the sequential DCT-based mode.

Lossless coding of JPEG is achieved by a simple predictive method, which combines up to three neighboring pixels (a, b, c) to form a prediction of the current pixel x to be encoded, as shown in Fig. 3.3. The prediction schemes of the lossless mode are presented in Table 3.1. No quantization is applied to the prediction of the coded pixel to achieve lossless coding. In the hierarchical mode, an image is encoded in a sequence of frames having different resolutions of the image, as shown in Fig. 3.4.

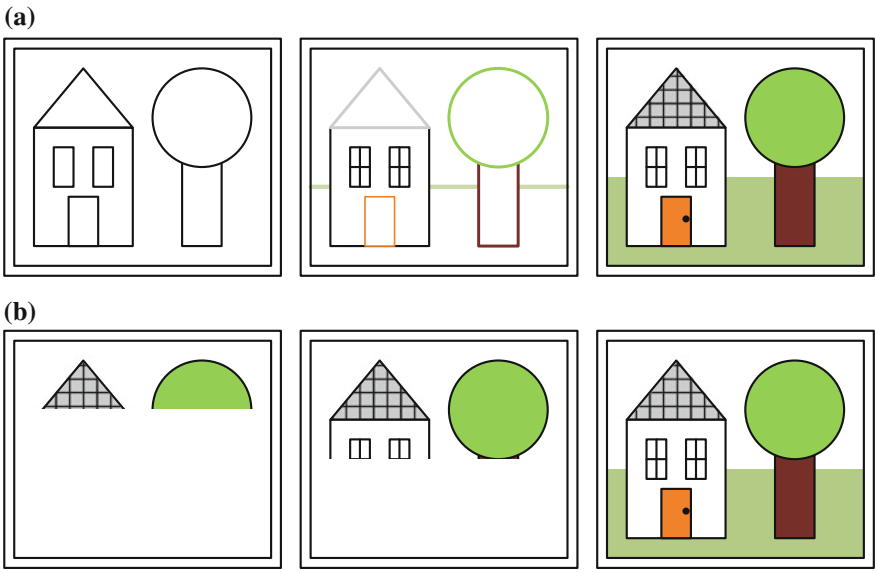

Fig. 3.2 Presentation of images using **a** progressive and **b** sequential coding

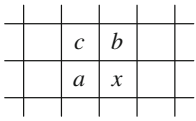

Fig. 3.3 Prediction scheme used in lossless mode of JPEG

Table 3.1 Prediction schemes for lossless mode of JPEG

Selection value	Prediction scheme
0	No prediction
1	$x = a$
2	$x = b$
3	$x = c$
4	$x = a + b - c$
5	$x = a + [0.5(b - c)]$
6	$x = b + [0.5(a - c)]$
7	$x = 0.5(a + b)$

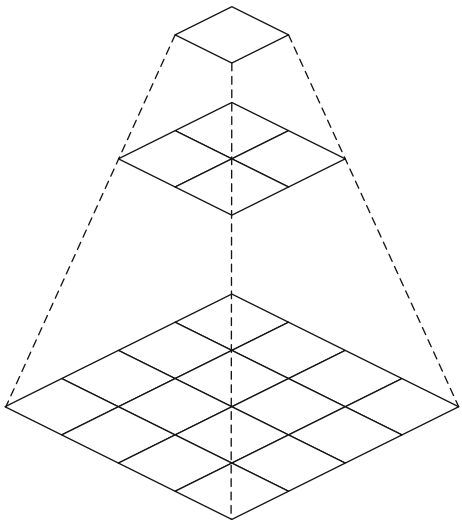

Fig. 3.4 Multi-resolution encoding of the hierarchical mode of JPEG

3.1.2 Baseline Sequential Codec of JPEG Standard

This section presents a brief introduction of the baseline sequential codec of the JPEG standard, which is the most popular mode of JPEG to date. Figure 3.5 show the major processing blocks of the sequential DCT encoder and decoder, respectively. The input image is first divided into 8 × 8 sub-images and the 8 × 8 DCT-II is applied to each sub-image. The $M \times N$ DCT-II and the inverse $M \times N$ DCT-II (IDCT-II) of the **n**th block are defined as

$$C(\mathbf{k}, \mathbf{n}) = \frac{2}{\sqrt{MN}} \alpha_{k_1} \alpha_{k_2} \sum_{p_1=0}^{M-1} \sum_{p_2=0}^{N-1} c(\mathbf{p}, \mathbf{n}) \cos\left(\frac{(2p_1 + 1)k_1\pi}{2M}\right) \cos\left(\frac{(2p_2 + 1)k_2\pi}{2N}\right), \tag{3.1}$$

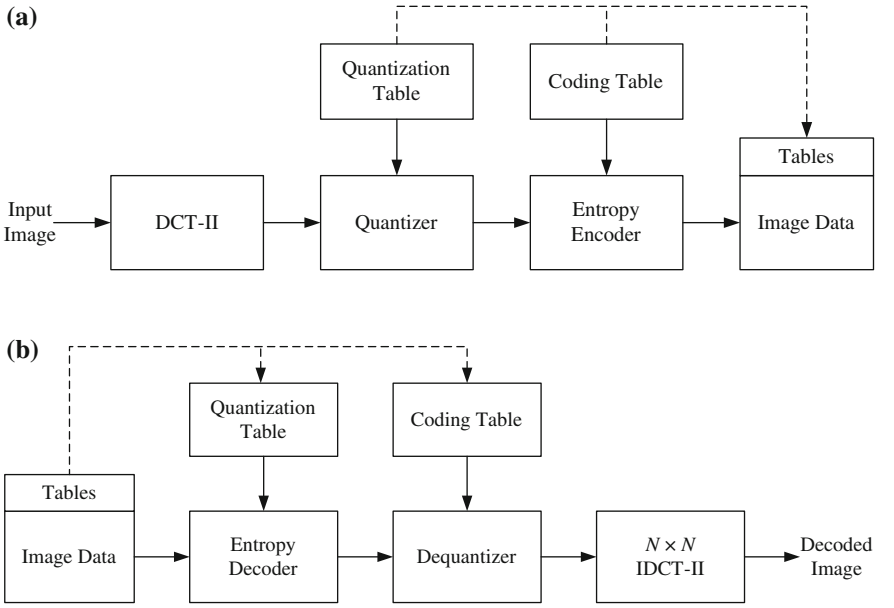

Fig. 3.5 Block diagrams of sequential DCT **a** encoder and **b** decoder

and

$$c(\mathbf{p},\mathbf{n}) = \frac{2}{\sqrt{MN}} \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} \alpha_{k_1} \alpha_{k_2} C(\mathbf{k},\mathbf{n}) \cos\left(\frac{(2p_1+1)k_1\pi}{2N}\right) \cos\left(\frac{(2p_2+1)k_2\pi}{2N}\right), \tag{3.2}$$

respectively, where

$$\alpha_u = \begin{cases} \frac{1}{\sqrt{2}}, & u = 0, \\ 1, & \text{otherwise}, \end{cases} \tag{3.3}$$

and $\mathbf{p} = (p_1, p_2)$. For the baseline sequential codec of the JPEG standard, M and N are chosen to be 8, and the DCT-II coefficient at $\mathbf{k} = (0,0)$ and the remaining 63 DCT-II coefficients are commonly referred as the "DC" and "AC" coefficients, respectively.

After the DCT-II operation, the 64 DCT-II coefficients are uniformly quantized by an 8 × 8 quantization matrix. Each element in the quantization matrix ranges from 1 to 255, and the quantized coefficient is given as

$$C'(\mathbf{k},\mathbf{n}) = \text{round}\left(\frac{C(\mathbf{k},\mathbf{n})}{Q(\mathbf{k})}\right), \tag{3.4}$$

where round(.) produces the nearest integer, $C'(\mathbf{k}, \mathbf{n})$ is the quantized DCT-II coefficients, and $Q(\mathbf{k})$ is the **k**th element of the quantization matrix. Dequantization is achieved by multiplying the quantized DCT-II coefficient $C'(\mathbf{k}, \mathbf{n})$ with $Q(\mathbf{k})$. Examples of the quantization matrices listed in Annex K of [ISO94] are shown in Tables 3.2 and 3.3.

These matrices are obtained from a series of psychophysical experiments conducted by Lohscheller [Loh84]. The visibility threshold of a DCT-II basis function superimposed on a uniform background is determined using

$$i_R = i_U + \gamma i_S, \tag{3.5}$$

where i_R, i_U, and i_S represent the resultant test image, image with uniform background, and visual stimulus (DCT-II basis function for this case), respectively. The visibility threshold of the visual stimulus is computed as the smallest value of γ to yield a JND between the uniform background and visual stimulus.

To reduce the bits required to encode the quantized DCT coefficients, the quantized DC and AC coefficients are entropy coded using Huffman coding. Huffman coding is a variable-length coding scheme that assigns shorter codewords to more frequently occurring symbols so that the average codeword length is close to entropy. To exploit the spatial correlation between DC coefficients of adjacent blocks, the quantized DC coefficient is encoded using differential encoding. The difference DIFF is computed by subtracting the current quantized DC coefficient

Table 3.2 Example of luminance quantization matrix from Annex K of [ISO94]

Luminance quantization matrix							
16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
29	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Table 3.3 Example of chrominance quantization matrix from Annex K of [ISO94]

Chrominance quantization matrix							
17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Table 3.4 Categories of differential values of DIFF

Category SSSS	Huffman codeword of SSSS	Differential values of DIFF
0	00	0
1	010	−1, 1
2	011	−3, −2, 2, 3
3	100	−7, …, −4, 4, …, 7
4	101	−15, …, −8, 8, …, 15
5	110	−31, …, −16, 16, …, 31
6	1110	−63, …, −32, 32, …, 63
7	11110	−127, …, −64, 64, …, 127
8	111110	−63, …, −128, 128, …, 255
9	1111110	−511, …, −256, 256, …, 511
10	11111110	−1023, …, −512, 32, …, 1023
11	111111110	−2047, …, −1024, 1024, …, 2047

with the previous DC coefficient. For a precision of 8 bits per sample, each sample is level-shifted to a signed representation by subtracting 128. Hence, the largest DC coefficient falls in the range [−1024, 1016] and the values of DIFF fall within [−2040, 2040].

The possible values of DIFF are grouped into 12 categories SSSS as shown in Table 3.4, and the codeword of the quantized DC coefficient is shown in Fig. 3.6. The codeword of each quantized DC coefficient is formed by concatenating the Huffman codeword of SSSS (except for SSSS = 0) and the two's complement of DIFF or DIFF −1. When DIFF is positive, the "SSSS" low-order bits of DIFF are added after the Huffman codeword of SSSS. When DIFF is negative, the "SSSS" low-order bits of DIFF −1 are added after the Huffman codeword of SSSS.

Prior to the encoding of quantized AC coefficients, the quantized AC coefficients are ordered into a 1-D sequence shown in Fig. 3.7. The key objective of this reordering scheme is to form long runs of '0's, which are introduced after quantization. The reordered sequence of the quantized AC coefficients is then run-length coded. Using run-length coding, up to 16 zero quantized AC coefficients followed by a non-zero AC coefficient can be encoded at any one time. For a precision of 8 bits per sample, the values of the quantized AC coefficient are grouped into 10 categories SSSS, as shown in Table 3.5. Together with the run-length RRRR, the symbol RRRRSSSS can be encoded into a Huffman codeword ranging from 2 to 16 bits using the example tables in Annex K of the JPEG standard, and the codeword

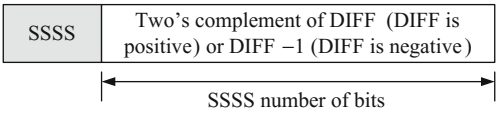

Fig. 3.6 Format of codeword of quantized DC coefficient

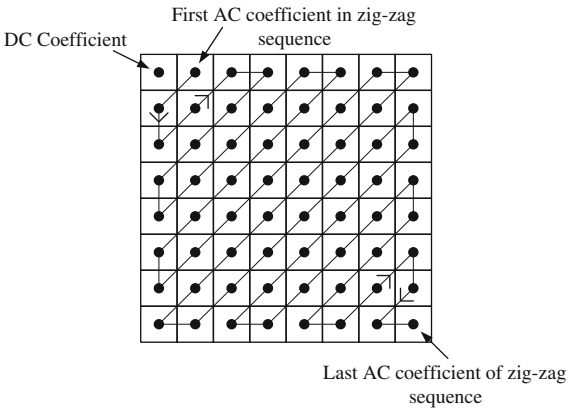

Fig. 3.7 Reordering of AC coefficients using the zig-zag sequence

Table 3.5 Categories of values of quantized AC coefficient

Category SSSS	Values of quantized AC coefficient
1	−1, 1
2	−3, −2, 2, 3
3	−7, …, −4, 4, …, 7
4	−15, …, −8, 8, …, 15
5	−31, …, −16, 16, …, 31
6	−63, …, −32, 32, …, 63
7	−127, …, −64, 64, …, 127
8	−63, …, −128, 128, …, 255
9	−511, …, −256, 256, …, 511
10	−1023, …, −512, 32, …, 1023

of the quantized AC coefficients (run-length and non-zero quantized AC coefficient) is formed by concatenating the Huffman codeword of RRRRSSSS and the two's complement of quantized AC coefficient, as shown in Fig. 3.8.

Two special symbols ZRL and EOB are used to represent a run-length of 16 zero coefficients and the end-of-block, respectively. In the case of a run-length of zero AC coefficients exceeding 16, multiple symbols of ZRL will be used. The symbol EOB is used when the remaining quantized AC coefficients to be encoded are zero.

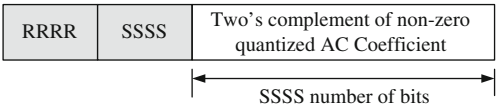

Fig. 3.8 Format of codeword of quantized AC coefficient

3.1.3 Blocking Artifact in JPEG Image

One of the obvious artifacts in heavily compressed JPEG images is blocking artifact. This artifact is characterized by the discontinuity between neighboring subimages [HL83], which is the consequence of coarse quantization of the AC coefficients. Generally, blocking artifact is highly visible in spatially smooth regions of an image. An example of the blocking artifact due to JPEG compression is shown in Fig. 3.9, and it is clear that the amount of blocking artifact in the hat, shoulder, and face regions of the "Lena" image increases as the bitrate reduces.

3.2 Computational Model for JND in DCT-II Domain

As seen in Fig. 3.9, the blocking artifact in JPEG compressed images is more prominent at low bitrates, and PICs are capable of achieving higher compression as compared to the standard JPEG encoder without significant loss of image quality.

The JPEG compatible PICs to be discussed in the following section shall be referred as Wat's [Wat93], Zhang's [ZLX05], and Wei's [WN09] coders. This class of PICs computes the JND threshold using (2.15), which estimates the luminance adaptation, contrast masking, and contrast sensitivity from the DCT-II coefficients. A generalized block diagram of these DCT-II based PICs is illustrated in Fig. 3.10.

3.2.1 Luminance Adaptation

As discussed in Chap. 2, the JND threshold of each DCT-II subband can be computed as a product of detection threshold and its elevation parameters given by luminance adaptation and contrast masking. The luminance adaptation that is applied in Watson's, Wei's, and Zhang's coders are plotted in Fig. 3.11.

Watson approximated the elevation parameter using a power function based on the formulation of the CSF proposed by Ahumada and Peterson [AP92]. Zhang et al. [ZLX05] approximated the elevation parameter using a square root equation for average intensity of DCT-II block lower and equal to 128 and a cube root equation for average intensity of DCT-II block higher than 128. On the other hand, Wei and Ngan [WN09] approximated the elevation parameter using two linear functions. The formulations proposed by Zhang et al. as well as Wei and Ngan suggest that HVS has lower sensitivities at low and high intensities, and these formulations are consistent with the findings from the experiments conducted in [SJ89, CL95, HK00]. The estimation of luminance masking based on Watson's formation may be the least inaccurate at lower intensities since it is the only formulation that suggests no elevation at zero average intensity.

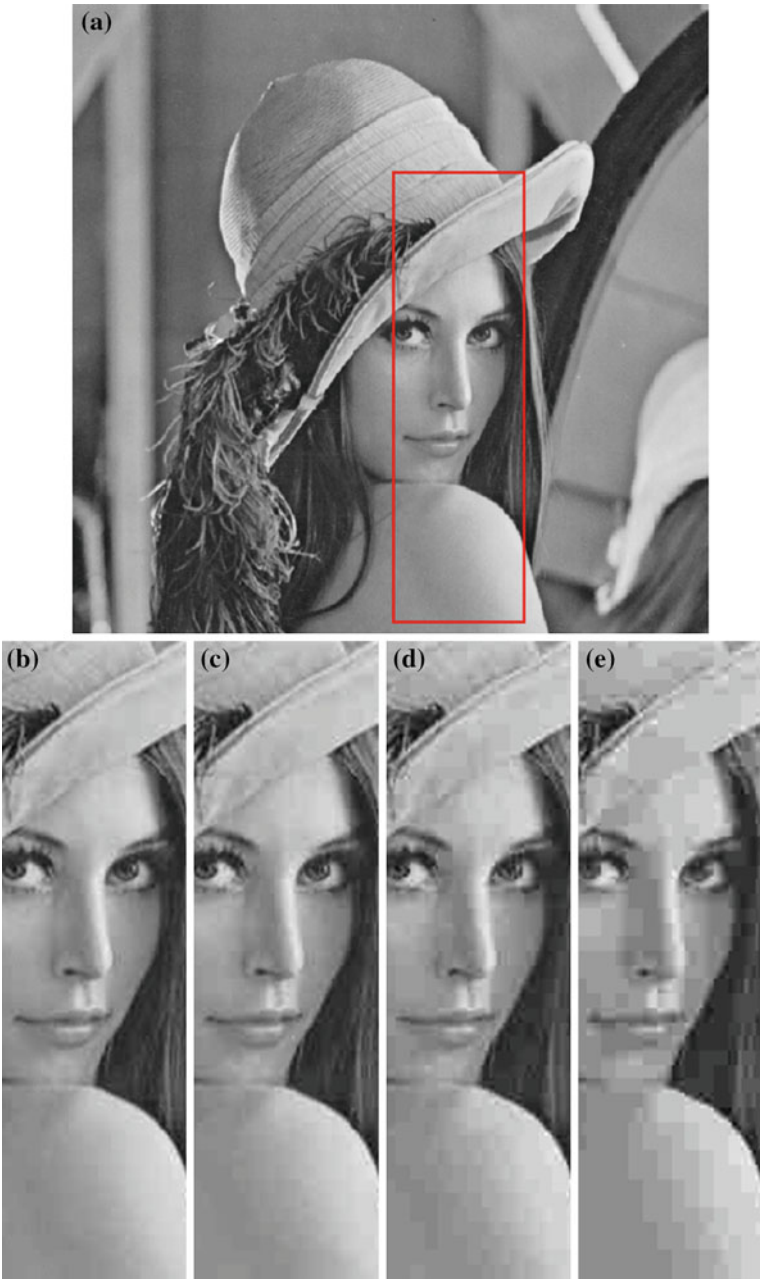

Fig. 3.9 Blocking artifact of JPEG compression with "Lena" image, where **a** is original image, and cropped part of "Lena" image (*in red box*) compressed at **b** 0.5 bpp, **c** 0.4 bpp, **d** 0.3 bpp, and **e** 0.2 bpp

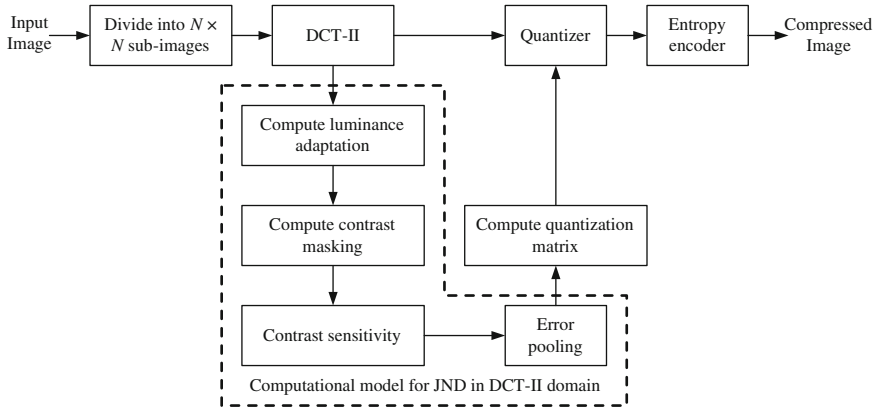

Fig. 3.10 Generalized block diagram of DCT based PIC with computational model for JND in DCT-II domain

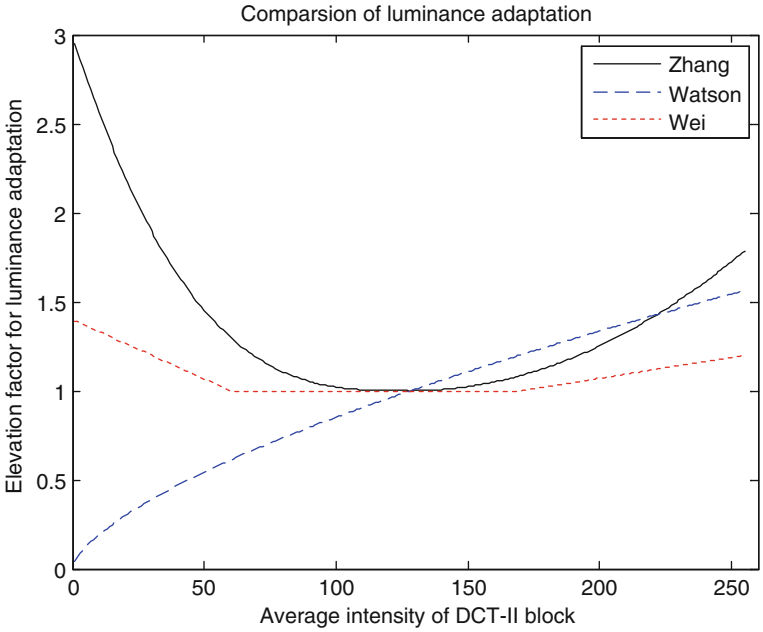

Fig. 3.11 Luminance adaptation of Zhang's, Watson's, and Wei's image coders

3.2.2 Block Classification

Zhang et al. [ZLX05, ZLX08] classify each $N \times N$ DCT-II block into plain, edge, and texture regions based on the energy of the DCT-II coefficients. While Zhang et al.'s approach is computational simple, their approach requires several thresholds

to be predetermined for each value of N. Their block classification results for $N = 8$ is reported in [ZLX05, ZLX08]. On the other hand, Wei and Ngan as well as Yang et al. used the Canny edge detector [Can86] to compute the edge map of an image, and this edge map is then used to locate the edge and texture regions of the image. While the Canny edge operator is very effective in detecting edges, this operator is highly computation intensive and is not designed to detect texture regions of image.

The block classification results of the "Barbara" and "Baboon" images from Wei's and Zhang's coders are shown in Fig. 3.12. It is observed that Wei's method produces more consistent results in the bookshelf and table regions of the "Barbara" image as compared to Zhang's method. However, Wei's method is found to be overly sensitive in the face, body, and pants regions of the "Barbara" image. This led to some plain areas in these regions to be detected as edges and textures. For the "Baboon" image, a large area of the image is classified as texture by Wei's method, which is consistent with human evaluation. On the other hand, some of these texture regions are classified into edges by Zhang's method. It is also interesting to note that Wei's method manages to detect some of the fine edges at the bottom left and right sides of the "Baboon" image, which are wrongly classified as plain region by Zhang's method.

3.3 Computational Model for JND in Pixel Domain

In this section, JPEG compatible PICs integrated with computational models for JND in the pixel domain shall be discussed. A generalized block diagram for this class of coders is illustrated in Fig. 3.13. Differing from the first class of PIC discussed in Sect. 3.2, the pixel-based JND model is converted into the DCT-II domain before it is integrated into the JPEG compatible PIC.

Luminance adaptation and contrast masking are generally used in computational models for JND in the pixel domain. For achromatic images, Yang et al. [YLL03, YLL05] derived a generalized expression for the spatial JND using NAMM. Let $t(\mathbf{x})$ denote the spatial JND threshold of the pixel located at $\mathbf{x}$, where $\mathbf{x} = \{x_1, x_2\}$, $x_1 = 0, 1, \ldots, H-1$, and $x_2 = 0, 1, \ldots, W-1$. Based on luminance adaptation and contrast masking, their spatial JND profile of an image is given as

$$t_{\mathrm{YLL}}(\mathbf{x}) = t_{la}(\mathbf{x}) + t_{cm}(\mathbf{x}) - C_{ol} \times \min\{t_{la}(\mathbf{x}), t_{cm}(\mathbf{x})\}, \tag{3.6}$$

where $t_{la}(\mathbf{x})$ and $t_{cm}(\mathbf{x})$ are the visibility thresholds due to luminance adaptation and contrast masking at $\mathbf{x}$, respectively; C_{ol} accounts for the reduction of spatial JND due to the overlapping effect in masking and $0 < C_{ol} \leq 1$.

The spatial JND profiles of the "Bicycle", "Pepper", and "Elaine" images computed using (3.6) are shown in Fig. 3.14. The lighter and darker regions of the JND profiles have higher and lower JND thresholds, respectively. Higher JND threshold due to contrast masking is found in the edge and texture regions of the

Fig. 3.12 Block classification results of "Barbara" and "Baboon" images shown in (**a**) and (**b**), respectively, using (**c**, **d**) Wei's method, and (**e**, **f**) Zhang's method. *Black*, *gray* and *white regions* indicate plain, texture and edge regions, respectively

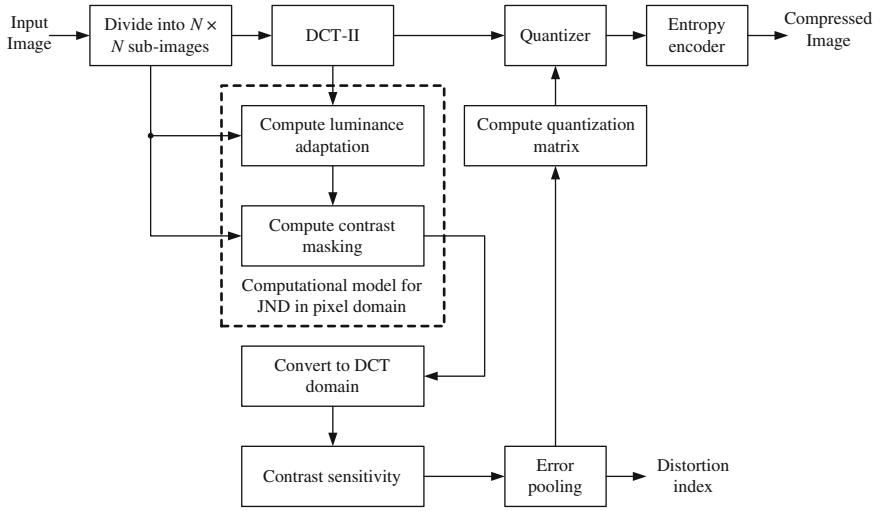

Fig. 3.13 Generalized block diagram of DCT based PIC with computational model for JND in pixel domain

images. In addition, higher JND threshold due to luminance adaptation is found in regions having low intensity.

The spatial profiles $t_{\mathrm{CL}}(\mathbf{x})$ and $t_{\mathrm{CB}}(\mathbf{x})$ computed by the pixel-based JND models in [CL95, CB99], respectively, are special cases of (3.6). Chou and Li proposed a pixel-based JND model that uses a simplified relationship between luminance adaptation and contrast masking [CL95, CC96]. The spatial JND computed by Chou and Li's JND model can be obtained using (3.6) by letting $C_{ol} = 1$, and we have

$$t_{\mathrm{CL}}(\mathbf{x}) = \max\{t_{la}(\mathbf{x}), t_{cm}(\mathbf{x})\}. \tag{3.7}$$

Chiu and Berger considered luminance adaptation to be the major contributor of their JND model [CB99]. By considering $\min\{t_{la}(\mathbf{x}), t_{cm}(\mathbf{x})\} \equiv t_{cm}(\mathbf{x})$, (3.6) becomes

$$t_{\mathrm{CB}}(\mathbf{x}) = t_{la}(\mathbf{x}) + (1 - C_{ol})t_{cm}(\mathbf{x}), \tag{3.8}$$

where $(1 - C_{ol})$ is experimentally found to be between 0.5 and 1.

3.3.1 Decomposition of Spatial JND Profile of an Image

Using a set of weights computed using Mannos and Sakrison's CSF [MS74], Chou and Li [CL95] decompose the spatial JND profile of an image into subband JND

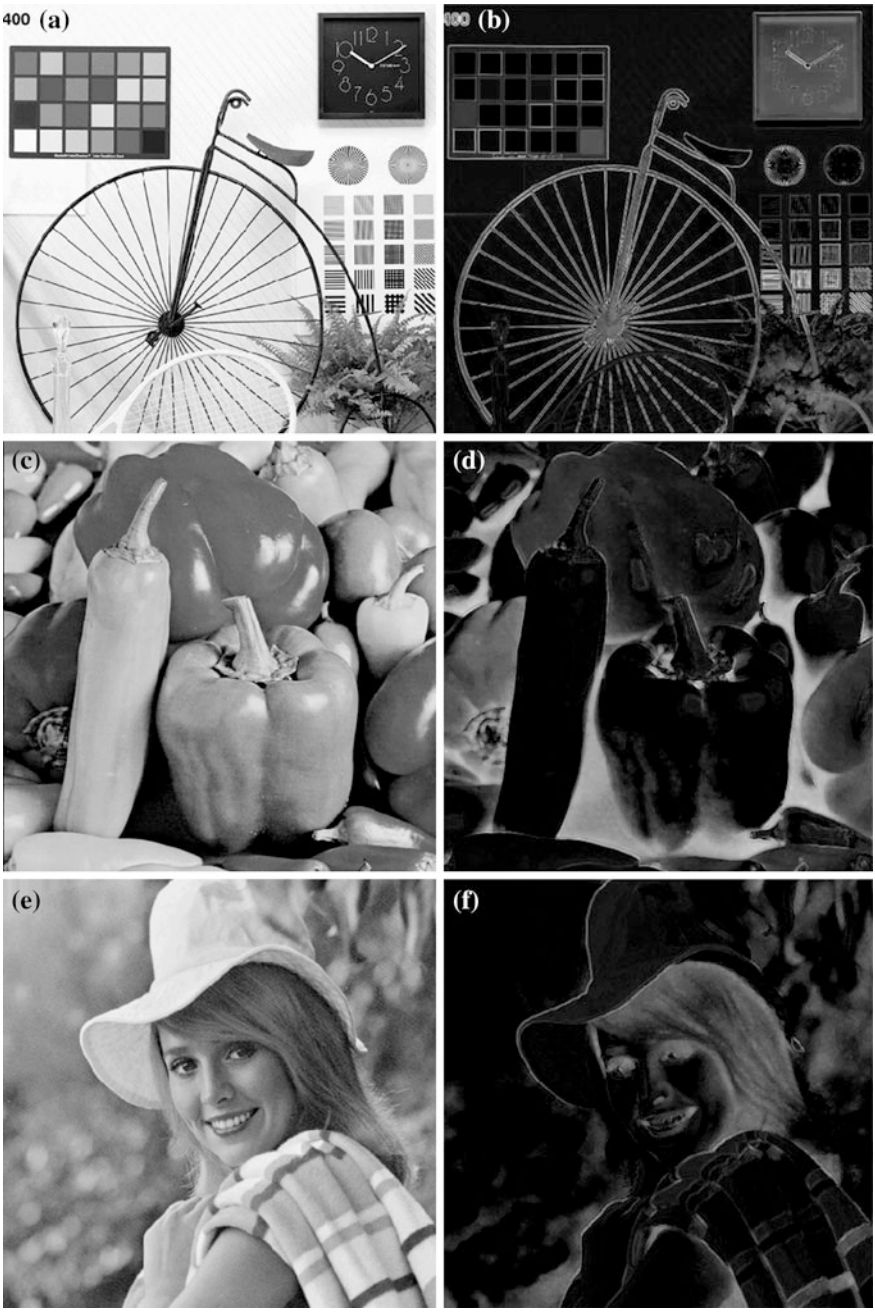

Fig. 3.14 Spatial JND profiles of "Bicycle", "Pepper", and "Elaine" images obtained with (3.6) are shown in (**b**), (**d**), and (**f**), respectively. "Bicycle", "Pepper", and "Elaine" images are shown in (a), (c), and (e), respectively. *Darker* and *brighter* regions indicate lower and higher JND values, respectively. These spatial JND profiles are contrast stretched to improve visibility

profiles. Assuming the spatial JND profile is decomposed into B subband JND profiles, let $t_B(b_i, \mathbf{s})$ denote the JND threshold of the b_i th subband JND profile at $\mathbf{s}$, where $b_i = 0, 1, \ldots, B-1$, $\mathbf{s} = (s_1, s_2)$, $s_1 = 0, 1, \ldots, \left(H/\sqrt{B}\right) - 1$, and $s_2 = 0, 1, \ldots, \left(W/\sqrt{B}\right) - 1$. The B subband JND profiles are estimated as

$$t_B(b_i, \mathbf{s}) = \sqrt{w_B(b_i) \sum_{p_1=0}^{\sqrt{B}-1} \sum_{p_2=0}^{\sqrt{B}-1} t^2\left(s_1\sqrt{B} + p_1, s_2\sqrt{B} + p_2\right)}, \tag{3.9}$$

where $w_B(b_i)$ is the weighting factor derived from a parametric CSF. The b_ith weighting factor computed by [CL95] is given by

$$w_B(b_i) = \left(d_B(b_i) \sum_{a_1=0}^{B-1} d_B^{-1}(a_1)\right)^{-1}, \tag{3.10}$$

and $d_B(b_i)$ is computed as

$$d_B(b_i) = \sum_{a_1=0}^{W/\sqrt{B}-1} \sum_{a_2=0}^{H/\sqrt{B}-1} CSF\left(\frac{W}{\sqrt{B}} \left\lfloor \frac{b_i}{\sqrt{B}} \right\rfloor + a_1, \frac{W}{\sqrt{B}} \bmod_{\sqrt{B}} (b_i) + a_2\right), \tag{3.11}$$

where $\bmod_b(a)$ produces the reminder of a divided by b, and CSF denotes a parametric CSF.

The parametric CSF [MS74, Nil85, NLS86, CR90, WN09] at $\mathbf{x}$ is expressed as

$$CSF(\mathbf{x}) = a_{\mathrm{CSF}}(b_{\mathrm{CSF}} + c_{\mathrm{CSF}} f(\mathbf{x})) \exp\left((-c_{\mathrm{CSF}} f(\mathbf{x}))^{d_{\mathrm{CSF}}}\right), \tag{3.12}$$

where $f(\mathbf{x})$ is the spatial frequency in cpd; a_{CSF}, b_{CSF}, c_{CSF}, and d_{CSF} are the parameters of the CSF. Assuming the HVS is isotropic [MS74], Mannos and Sakrison [MS74] empirically obtained the CSF as

$$CSF_{\mathrm{MS}}(\mathbf{x}) = 2.6(0.0192 + 0.114 f(\mathbf{x})) \exp\left(-(0.114 f(\mathbf{x}))^{1.1}\right), \tag{3.13}$$

and the spatial frequency is expressed as $f(\mathbf{x}) = \sqrt{(x_1/v_{x_1})^2 + (x_2/v_{x_2})^2}$. Let Λ_{x_1} and Λ_{x_2} denote the height and width of a pixel, respectively, and the visual angles are

$$\begin{aligned} v_{x_1} &= 2\mathrm{a}\tan(\Lambda_{x_1}/2d), \\ v_{x_2} &= 2\mathrm{a}\tan(\Lambda_{x_2}/2d). \end{aligned} \tag{3.14}$$

If the viewing distance d is defined as six times of the image height [ITU02], the vertical visual angle v_{x_1} is 0.0187°. Assuming the pixels of the display are square, we then have 53.76 pixels subtended 1° vertically and horizontally.

3.3.2 Parametric CSF

The ModelFest data set is a collection of experimental data from 10 labs that is used to test and calibrate models for spatial contrast detection [WA05]. The stimuli used in the ModelFest experiments subtend a viewing angle of 2.133° (falls within the foveal region), and the viewing of the stimuli is in binocular vision (both eyes are used to view stimuli). As CSF remains fairly constant for wide visual angle up to 120° [NAW93], we can refer to the observations of ModelFest in our comparative analysis of CSFs at larger viewing angles.

Nill [Nil85] pointed out that the parametric CSFs obtained using sine or square grating functions may not lead to good estimation of the HVS for DCT-II based applications since DCT-II implies even extension of the input sequence. To avoid this issue, Nill proposed to scale the CSF using

$$|A(f)| = \sqrt{\frac{1}{4} + \frac{1}{\pi^2}\left[\ln\left(\left(\frac{2\pi}{\alpha}f\right) + \sqrt{\frac{4\pi^2}{\alpha^2}f^2 + 1}\right)\right]^2}, \tag{3.15}$$

where

$$\alpha^{-1} = 2a\sin\left(\frac{1.5}{2\sqrt{0.5625 + d^2}}\right). \tag{3.16}$$

The CSFs from Mannos and Sakrison [MS74], weighted CSF from Mannos and Sakrison, Nill [Nil85], Ngan et al. [NLS86], Chitprasert and Rao [CR90], as well as Wei and Ngan [WN09] are plotted in Fig. 3.15. These CSFs are referred as CSF_{MS}, CSF_{Nil}, CSF_{NLS}, CSF_{CR}, and CSF_{WN}, respectively. The empirically obtained CSFs proposed by Chitprasert and Rao as well as Wei and Ngan are optimized for DCT-II, whereas Mannos and Sakrison computed their CSF using sine grating functions. One additional CSF referred as CSF_{WMS} is obtained by scaling CSF_{MS} using (3.15). The parameters of the five CSFs are summarized in Table 3.1. The spatial frequency of these CSF is computed by [CR90]

$$f_{\mathrm{DCT}}(\mathbf{k}) = \frac{1}{2N}\sqrt{(k_1/v_{x_1})^2 + (k_2/v_{x_2})^2}. \tag{3.17}$$

For $N = 8$, the peak responses of CSF_{MS}, CSF_{WMS}, CSF_{Nil}, CSF_{NLS}, CSF_{CR}, and CSF_{WN} are found to be 6.72, 10.08, 6.72, 3.36, 3.36, and 0 cpd, respectively.

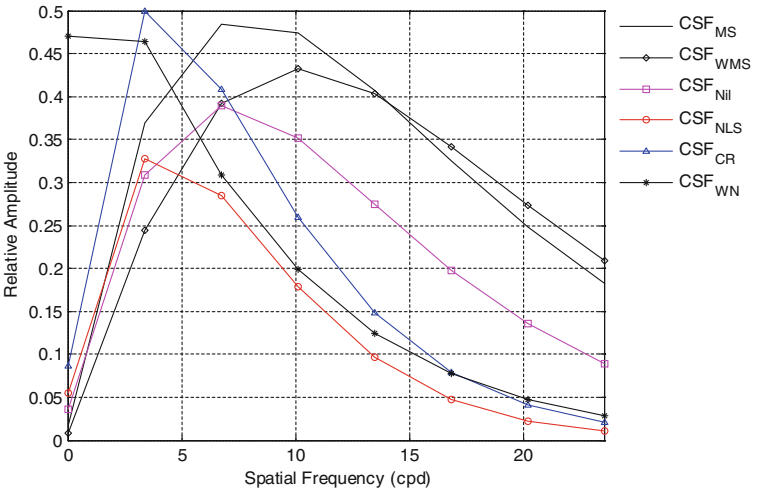

Fig. 3.15 Comparison of six parametric CSFs for $N = 8$

It is interesting to note that only CSF_{WN} estimated a slight increase in response as the spatial frequency reduces from 3.36 to 0 cpd. This observation differs from the one made by Watson and Ahumada [WA05] using the ModelFest data. In fact, Watson and Ahumada observed that the CSF peaks around 3–4 cpd and such response is also found in CSF_{NLS} and CSF_{CR}. However, the response of CSF_{NLS} is the lowest among the six CSFs and leads to the highest detection threshold (inversely proportional to contrast sensitivity) which may be undesirable in some applications.

3.4 Computing Quantization Matrix

As highlighted in [WA05], the accuracy of the estimation for CSF can be further enhanced by considering the oblique effect [BKW75, PW84, AP92] and the spatial summation effect [PAW93a]. The oblique effect of two DCT basis vectors is first discussed in [AP02], which involves scaling the first component in the RHS of (2.8) with the term $[r + (1 - r)\cos^2\theta(\mathbf{k})]$. Considering the oblique and spatial summation effects, the base detection threshold for DCT-II basis becomes [WN09]:

$$T_{b,CSF}(\mathbf{k}) = \frac{s\,\exp\left((c_{\mathrm{CSF}} f(\mathbf{k}))^{d_{\mathrm{CSF}}}\right)}{\alpha_{k_1}\alpha_{k_2} a_{\mathrm{CSF}}(b_{\mathrm{CSF}} + c_{\mathrm{CSF}} f(\mathbf{k}))[r + (1 - r)\cos^2\theta(\mathbf{k})]}, \tag{3.18}$$

where $s = 0.25$ accounts for the spatial summation effect [PAW93a]. Let $q(\mathbf{k})$ denote the elements of the quantization matrix. Since the maximum quantization error is half of the quantization step, the quantization matrix can be computed as

$$q(\mathbf{k}) = \frac{2MT_{b,CSF}(\mathbf{k})}{L_{\max} - L_{\min}}. \tag{3.19}$$

3.4.1 Computing Quantization Matrix with Spatial JND Profile

Chou and Li [CL95] used a 2-D quadrature mirror filterbank (QMF) [Joh80] to divide the input image into 16 subbands. To determine perceptually significant coefficients in each subband, the spatial JND profile of the input image is also decomposed into 16 subband JND profiles.

Referring to (3.1), DCT-II can be considered as a filterbank that divides each 8×8 sub-image into 64 DCT subbands [MSO07], and the perceptual importance of the 64 DCT subbands can be determined using 64 subband JND profiles. Let $w_{64}(k_q)$ denote the weighting factor of the 64 subband JND profiles, where $k_q = 0, 1, \ldots, 63$; $w(u, v)$ denote the elements of weighting matrix $\mathbf{W}$; u and v denote the row and column of $\mathbf{W}$, respectively. The elements of $\mathbf{W}$ are determined as $w(u, v) = w_{64}(8u + v)$. Using (3.10), $\mathbf{W}$ is found to be

$$\mathbf{W} = \begin{bmatrix} 0.0078 & 0.0051 & 0.0051 & 0.0061 & 0.0081 & 0.0115 & 0.0169 & 0.0257 \\ 0.0055 & 0.0050 & 0.0053 & 0.0064 & 0.0085 & 0.0119 & 0.0175 & 0.0265 \\ 0.0049 & 0.0051 & 0.0056 & 0.0069 & 0.0092 & 0.0128 & 0.0187 & 0.0282 \\ 0.0052 & 0.0055 & 0.0063 & 0.0078 & 0.0103 & 0.0143 & 0.0207 & 0.0311 \\ 0.0060 & 0.0064 & 0.0074 & 0.0091 & 0.0120 & 0.0165 & 0.0237 & 0.0352 \\ 0.0073 & 0.0078 & 0.0090 & 0.0111 & 0.0144 & 0.0196 & 0.0279 & 0.0410 \\ 0.0093 & 0.0099 & 0.0113 & 0.0138 & 0.0178 & 0.0240 & 0.0337 & 0.0490 \\ 0.0121 & 0.0129 & 0.0147 & 0.0177 & 0.0255 & 0.0300 & 0.0417 & 0.0599 \end{bmatrix}. \tag{3.20}$$

Let $q(u, v)$ denote the elements of the quantization matrix, and the elements of the quantization matrix are defined as [TG11]

$$q(u, v) = 2\sqrt{w(u, v)} \times \min_{(n_1, n_2)} \left\{ \sqrt{\sum_{e=0}^{7} \sum_{f=0}^{7} t^2(8n_1 + e, 8n_2 + f)} \right\}, \tag{3.21}$$

where $n_1 = 0, 1, \ldots, H/N - 1$ and $n_2 = 0, 1, \ldots, W/N - 1$.

3.5 Summary

This chapter first presents an overview of the JPEG standard, with the focus on the sequential DCT mode. Two possible extensions of the sequential coder using JND models that are based on pixel and DCT-II subband are then discussed. These models consider the effects of contrast sensitivity, luminance adaptation, and contrast masking.

We compared the luminance adaptation proposed by Watson [Wat93], Zhang et al. [ZLX05], as well as Wei and Ngan [WN09]. As compared to Watson's implementation, Zhang's and Wei's implementations are found to be in closer agreement with the HVS since these implementations account for lower sensitivity at lower and higher intensities.

Subsequently, two approaches for block classification are discussed in this chapter. Zhang et al. proposed an approach in the DCT-II domain that classifies each 8×8 sub-image into plain, edge, and texture regions. On the other hand, Wei and Ngan's approach involves computing the average number of edge pixels in each $N \times N$ sub-image. Our comparison of these techniques revealed that the classification in the pixel-domain is superior in detecting texture regions as well as fine edges in the test-images.

Chou and Li [CL95] used the parametric CSF proposed by Mannos and Sakrison [MS74] to decompose the spatial JND profile into subband JND profiles. A comparative analysis of several parametric CSFs [MS74, Nil85, NLS86, CR90, WN09] is presented in Sect. 3.3.2. While these parametric CSFs perform similar, we found that the CSFs proposed by Ngan et al. [NLS86] as well as Chitprasert and Rao [CR90] exhibit similar responses to those reported in ModelFest [WA05].

Since CSF is usually defined in the spatial frequency domain, many JND models in the pixel domain [CL95, CC96, CB99, YLL03, YLL05, LLP10] do not account for contrast sensitivity. One possible workaround for this limitation is to convert the JND model from the pixel to the subband domain [CL95], and then integrate the CSF into the converted JND model in the subband domain [TG11].

In the next chapter, we validate JND model in pixel and subband domains using subjective experiments. A comparative analysis of these JND models and their performance in image compression shall be presented.

Chapter 4
Validation of Computational Model for JND

The limitations of HVS prevent it from sensing all changes in a reconstructed image after compression. By exploiting these limitations of the HVS, PICs are able to achieve higher compression with lesser visual degradation as compared to non-PICs. Since the performance of PIC is largely dependent on the estimation accuracy of visual degradation using computational model for JND, this chapter focuses on the validation of such computational models using a series of subjective experiments.

This chapter starts with a discussion of a technique that facilitates comparative analysis of JND models that are computed from subbands and pixels of an image [ZLX08]. This technique involves estimating the spatial JND profile using the JND profile computed in the DCT subband domain, and the comparison of spatial JND profiles are then performed in the pixel domain. In this chapter, the noise-shaping performance of five JND models in the subband and pixel domains are studied.

Subjective experiments are conducted to examine the visibility of noise in noise-contaminated images, which are created using the spatial JND profile of images. The key differences between spatial JND profiles that are converted from subbands and those directly computed from pixels of image, as well as their advantages and disadvantages, are presented in this chapter. The contrast sensitivity estimated by the JND models is also compared. This is carried out by adding two grating functions into some test-images and then examining their spatial JND profiles. Finally, the correlation between JND model and human perception of visual degradation is studied using the Pearson correlation and Spearman rank-order correlation, based on the guidelines detailed in the report [VQE03] from video quality experts group (VQEG).

The rest of this chapter is organized as follows. The technique to estimate the spatial JND profile from the JND profile computed in DCT subband domain is discussed in Sect. 4.1. This is followed by a comparison of spatial JND profiles of images estimated from the DCT-II domain and computed directly from the pixel domain. A comparative analysis of noise-shaping performance of Chou's, Yang's, Watson's, Wei's, and Zhang's JND models is presented in Sect. 4.2. This is followed by an analysis of contrast sensitivity estimated by the five JND models.

E.-L. Tan and W.-S. Gan, *Perceptual Image Coding with Discrete Cosine Transform*, SpringerBriefs in Signal Processing,
DOI 10.1007/978-981-287-543-3_4

Section 4.3 presents a performance analysis of the five JND models by comparing their estimation accuracy of visual degradation in compressed images. Finally, this chapter is summarized in Sect. 4.4.

4.1 Verification of JND Modeling

To compare JND models in the pixel and DCT subband domains, Zhang et al. [ZLX08] devised a technique which involves estimating the JND values of a pixel at $\mathbf{x}$ by summing the contribution of N^2 JND thresholds within the $\mathbf{n}$th DCT-II block.

Before estimating the spatial JND profile, Zhang et al. replaced the $\mathbf{k}$th DCT-II coefficients having smaller magnitude than JND threshold $T(\mathbf{k},\mathbf{n})$ with zero since these coefficients do not contribute to the spatial JND profile. Based on this idea, Zhang et al. defined a new set of JND threshold $T'(\mathbf{k},\mathbf{n})$ as

$$T'(\mathbf{k},\mathbf{n}) = \begin{cases} \text{sign}(C(\mathbf{k},\mathbf{n}))T(\mathbf{k},\mathbf{n}), & \text{for } |C(\mathbf{k},\mathbf{n})| \geq t(\mathbf{k},\mathbf{n}), \\ 0, & \text{otherwise}, \end{cases} \tag{4.1}$$

where $\text{sign}(C(\mathbf{k},\mathbf{n}))$ produces the sign of $C(\mathbf{k},\mathbf{n})$. The $\text{sign}(C(\mathbf{k},\mathbf{n}))$ operation is used to avoid discontinuity due to the zeroed JND threshold at $(\mathbf{k},\mathbf{n})$. Subsequently, the spatial JND $t'(\mathbf{x})$ in the pixel domain is estimated by summing all the JND thresholds at the $\mathbf{n}$th block using

$$t'(\mathbf{x}) = \sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1} \alpha_{k_1}\alpha_{k_2} T'(\mathbf{k},\mathbf{n}) \cos\left(\frac{(2p_1+1)k_1\pi}{2N}\right)\cos\left(\frac{(2p_2+1)k_2\pi}{2N}\right), \tag{4.2}$$

for $x_1 = n_1 N + p_1$, $x_2 = n_2 N + p_2$, $\mathbf{x} = \{x_1, x_2\}$, and $p_1, p_2 = 0, 1, \ldots N-1$.

4.1.1 Comparison of Spatial JND Profile

In this sub-section, we shall present a comparative analysis of Chou's [CL95], Yang's [YLL05], Watson's [Wat93], Wei's [WN09], and Zhang's [ZLX05] JND models. The spatial JND profile of an image from Chou's, Yang's, Watson's, Wei's, and Zhang's JND models shall be referred as t_{CL}, t_{YLL}, t'_{Wat}, t'_{WN}, and t'_{ZLX}, respectively, where t_{CL} and t_{YLL} are computed in the pixel domain and t'_{Wat}, t'_{ZLX}, and t'_{WN} are computed in the DCT-II domain and then converted to the pixel domain.

The spatial JND profiles of the "Lena" test-image are shown in Fig. 4.1, and the bright and dark regions of the spatial JND profile indicate regions having high and low JND thresholds, respectively. The spatial JND profiles from the five JND models indicate high JND threshold are found in the feather and hair regions of the

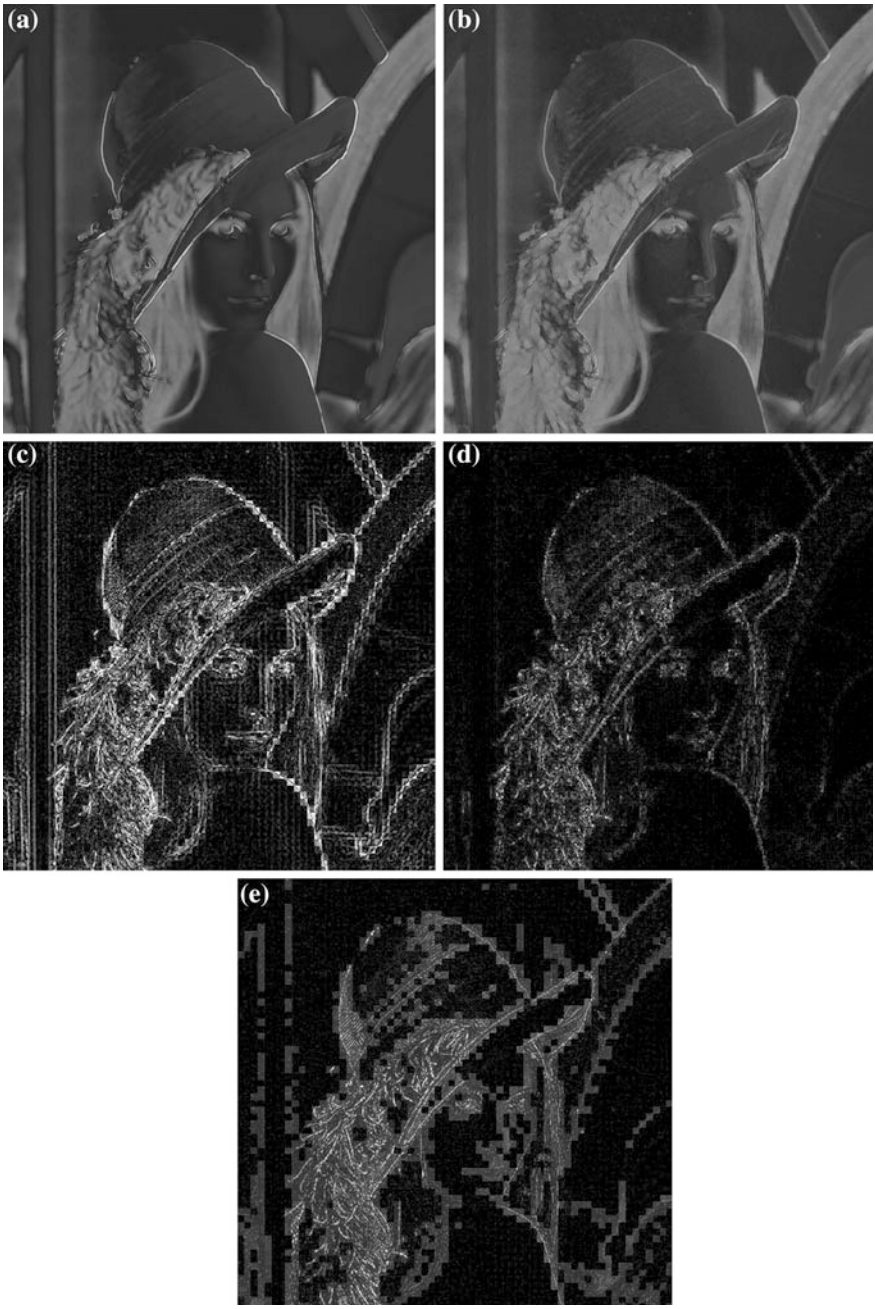

Fig. 4.1 Spatial JND profiles of "Lena" test-image obtained using Chou's, Yang's, Watson's, Wei's, and Zhang's JND models are shown in (**a**–**e**), respectively

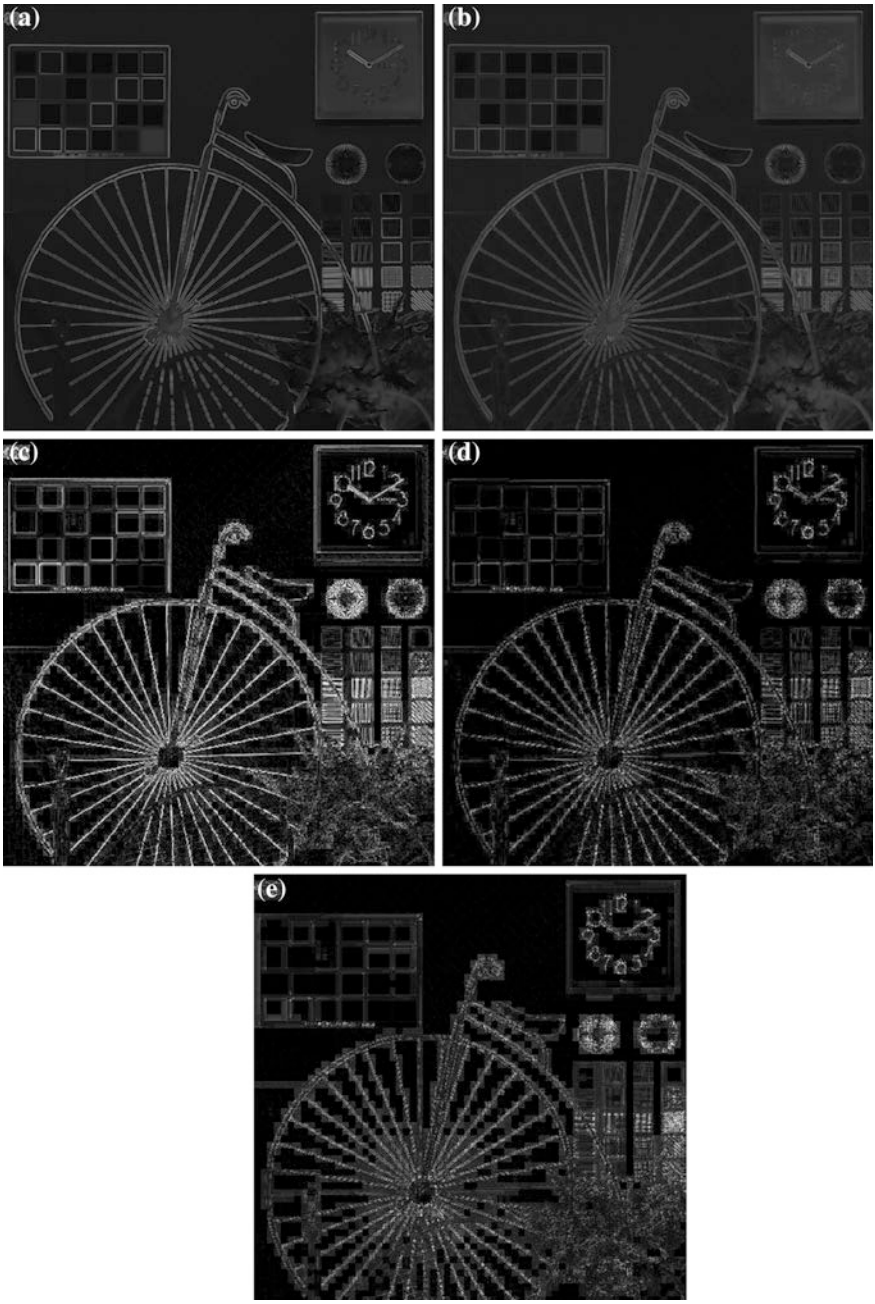

Fig. 4.2 Spatial JND profiles of "Bicycle" image obtained using Chou's, Yang's, Watson's, Wei's, and Zhang's JND models are shown in (**a**–**e**), respectively

"Lena" test-image. The spatial JND profiles t_{CL} and t_{YLL} are very similar, except t_{YLL} exhibits slightly lower JND threshold at the edges of the shoulder and hat regions. This is due to the fact that Yang's JND model estimates lower contrast masking for edges than textures whereas the same amount of contrast masking is estimated for edges and textures in Chou's JND model. Significant differences are found in the spatial JND profiles t'_{Wat}, t'_{WN}, and t'_{ZLX}, which are attributed to the different constrast masking used to compute these JND profiles. For instance, numerous bright blocks are found in the spatial JND profiles t'_{Wat} and t'_{ZLX}, especially at the edges of the shoulder and the hat regions of the "Lena" test-image. The changes of JND threshold in the spatial JND profile t'_{Wat} is the most abrupt among the three DCT-II based JND models. Such abrupt changes might lead to perceptible distortion due to large changes in the quantization matrices computed from these JND thresholds. On the contrary, the changes of the JND threshold in these regions are relatively gradual in t'_{WN}, and this spatial JND profile does not exhibit the block-like characteristic that is found in the spatial JND profiles t'_{Wat} and t'_{ZLX}. However, numerous small bright regions are found in the feather and hair regions of t'_{WN}, as seen in Fig. 4.1d.

The five spatial JND profiles of the "Bicycle" test-image are shown in Fig. 4.2. Similar observations are found in the "Lena" and "Bicycle" test-images, bright blocks are found at the edges of the bicycle, picture chart, and clock in the spatial JND profiles t'_{Wat} and t'_{ZLX}. Abrupt changes of the JND threshold are found in the spatial JND profiles t'_{Wat} and t'_{ZLX}, and more prominent in t'_{Wat}.

4.2 Noise-Shaping Performance of JND Model

To study the noise-shaping performance of a JND model, Chou and Li [CL95] compared the visual quality between noise-contaminated and original images. These images should be visually similar if the spatial JND profile of the image correlates with human perception. Based on the spatial JND profile $t(\mathbf{x})$ of an image, the noise-contaminated image is computed as

$$c_{nc}(\mathbf{x}) = c(\mathbf{x}) + \tau\,\mathrm{rand}(\mathbf{x})t(\mathbf{x}), \tag{4.3}$$

where $\mathrm{rand}(\mathbf{x})$ takes on values of 1 or -1 randomly; $c(\mathbf{x})$ and $c_{nc}(\mathbf{x})$ denote the original and noise-contaminated images, respectively; τ is a scaling factor larger than one to control the amount of error energy injected into the original image.

A comparison of the noise-shaping performance of five JND models is carried out using subjective experiments. Ten subjects made up of two females and eight males aged between 20 and 36 took part in the subjective experiments. At any one time, two images were displayed on the screen where the left and right images are the original and the noise-contaminated images, respectively. The subjects were

Table 4.1 Description of subjective score used in subjective evaluation

Subjective score	Description
0	Right image is the same quality as the left image
1	Right image is of very good quality as compared to left image, but some differences can be observed
2	Right image is worse than left image
3	Right image is much worse than left image

asked to record his or her assessment of each noise-contaminated image as compared to the original image according to Table 4.1.

Noise is injected into 20 test-images (see Fig. 4.3) from [URL01] to produce noise-contaminated images [CL95, YLL05, ZLX08]. The scaling factor τ in (4.1) is adjusted to produce noise-contaminated images having PSNR of 30 dB. An Acer P205H LCD monitor is used in this subjective experiment, and this monitor has a native resolution of 1600 × 900 pixels. The viewable screen of this monitor is 442.800 mm by 249.075 mm, therefore the width and the height of each pixel on this monitor are 0.277 mm. Since the images are of 512 × 512 pixels, the viewing distance (six times of the image height) is selected to be 850 mm.

The 95 % confidence interval of the subjective score for the contaminated images produced by the five JND models is shown in Fig. 4.4. The noise-contaminated images that are most similar to the original images are produced by t'_{WN}, and the contaminated images having the worst subjective score are produced by t_{CL}. The three DCT subband-based JND models outperform the two pixel-based JND models, and Zhang et al. [ZLX08] attributed the inferior performance of the pixel-based JND models to the omission of CSF. Despite numerous differences in the DCT subband-based JND models t'_{Wat}, t'_{WN}, and t'_{ZLX}, these JND models perform quite similarly.

Next, we inspect the performance of the five JND models in the presence of grain noise. The test-images selected for this experiment are TI-8 and TI-20, and the noise-contaminated images of TI-8 of TI-20 are shown in Figs. 4.5 and 4.6, respectively. We observed that the noise-contaminated images computed by the DCT subband-based JND models are comparably worse than those obtained using pixel-based JND models. Noise-contaminated images from DCT subband-based JND models exhibit more noise in the background, hair, and, eyes regions of TI-8 as compared to the noise-contaminated images obtained with pixel-based JND models. Specifically, the noise-contaminated images obtained using t'_{Wat}, t'_{WN}, and t'_{ZLX} exhibit visible distortion at the edges at the hair and left eye regions, especially at bright regions of TI-8. On the other hand, the background of the noise-contaminated images of TI-8 based on the spatial JND profiles t_{CL} and t_{YLL} contains significant amount of noise. The inferior performance of subband-based JND models might be attributed to the summing of the JND thresholds within an $N \times N$ block to estimate the JND threshold of a pixel, which prevents subband-based JND models to be highly accurate in estimating the JND threshold of a pixel in the presence of grain noise.

Fig. 4.3 Test-images used in our subjective experiments. Test-images from **a** to **t** are referred as TI-1 to TI-20, respectively

Figure 4.6 reveals highly perceptible distortion at the vertical stabilizer of the aircraft tail and the aircraft body for the noise-contaminated images obtained with the subband-based JND models, with Wei's JND model producing the least amount of perceptible distortion in these regions. Similar to TI-8 (see Fig. 4.5), the noise-contaminated images obtained with the pixel-based JND models have significant noise in the background of TI-20. There is significantly less distortion at the vertical

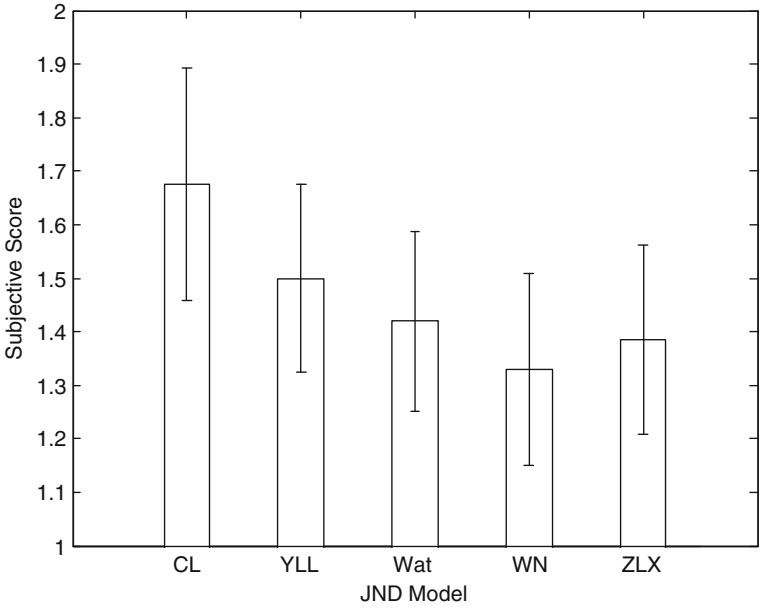

Fig. 4.4 Subjective score and 95 % confidence interval for contaminated images (PSNR = 30 dB) from five JND models

stabilizer of the aircraft tail and the aircraft body from the noise-contaminated images obtained with the pixel-based JND models, but the reduced distortion at these regions might be due to the under-estimation of JND thresholds with the pixel-based JND models.

4.2.1 Contrast Sensitivity Estimation with JND Model

Zhang et al. [ZLX08] concluded that the pixel-based JND models are inferior to the DCT subband-based JND models as pixel-based JND models do not consider the contrast sensitivity of the HVS. Adapting the approach used in [LMK98], Zhang et al. used two vertical gratings to evaluate the performance of five JND models in discerning noise (vertical grating injected into test-images as noise). Since the viewing distance used in our subjective experiments depends on the image height, we shall use horizontal grating in our subjective experiment and the grating frequencies are selected to be 4 and 20 cpd.

The five spatial JND profiles of TI-6 with horizontal gratings at 4 and 20 cpd are shown in Figs. 4.7 and 4.8, respectively. Since only the horizontal grating at 4 cpd falls within the sensitive range of the HVS, we expected this horizontal grating to

Fig. 4.5 Noise-contaminated images of TI-8. Original image of TI-8 is shown in (**a**), and the noise-contaminated images obtained with Chou's, Yang's, Watson's, Wei's, and Zhang's JND model are shown in (**b**–**f**), respectively

be highly visible and the horizontal grating at 20 cpd to be barely visible. Two observations can be drawn from the five spatial JND profiles of the test-images with the horizontal grating. First, pixel-based JND models are inconsistent with the HVS

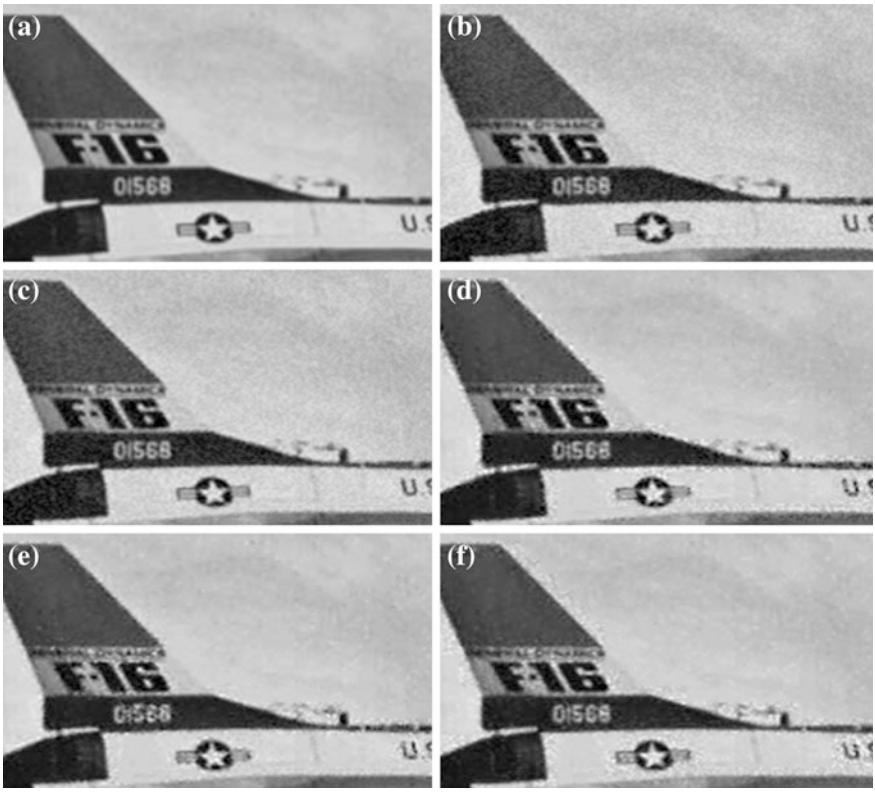

Fig. 4.6 Noise-contaminated images of TI-20. Original image of TI-20 is shown in (**a**), and the noise-contaminated images obtained with Chou's, Yang's, Watson's, Wei's, and Zhang's JND model are shown in (**b**–**f**), respectively

as the spatial JND profiles from these models do not reveal the highly visible grating at 4 cpd. The horizontal grating at 20 cpd is also undetected by the pixel-based JND models. Second, the DCT subband-based JND models are consistent with the HVS as only the horizontal grating at 4 cpd is detected by these JND models.

In some test-images, the two horizontal gratings are invisible to the HVS. This observation is generally found with images containing significant amount of high spatial frequency content, which leads to significant contrast masking and renders the horizontal gratings to be invisible. In such test-images, an accurate JND model should not detect both horizontal gratings at 4 and 20 cpd. The spatial JND profiles of TI-10 with horizontal gratings at 4 and 20 cpd are shown in Fig. 4.9. The spatial JND profile of TI-10 from each JND model with the horizontal gratings at 4 and 20 cpd are shown on the left and right, respectively. Consistent with the HVS, the

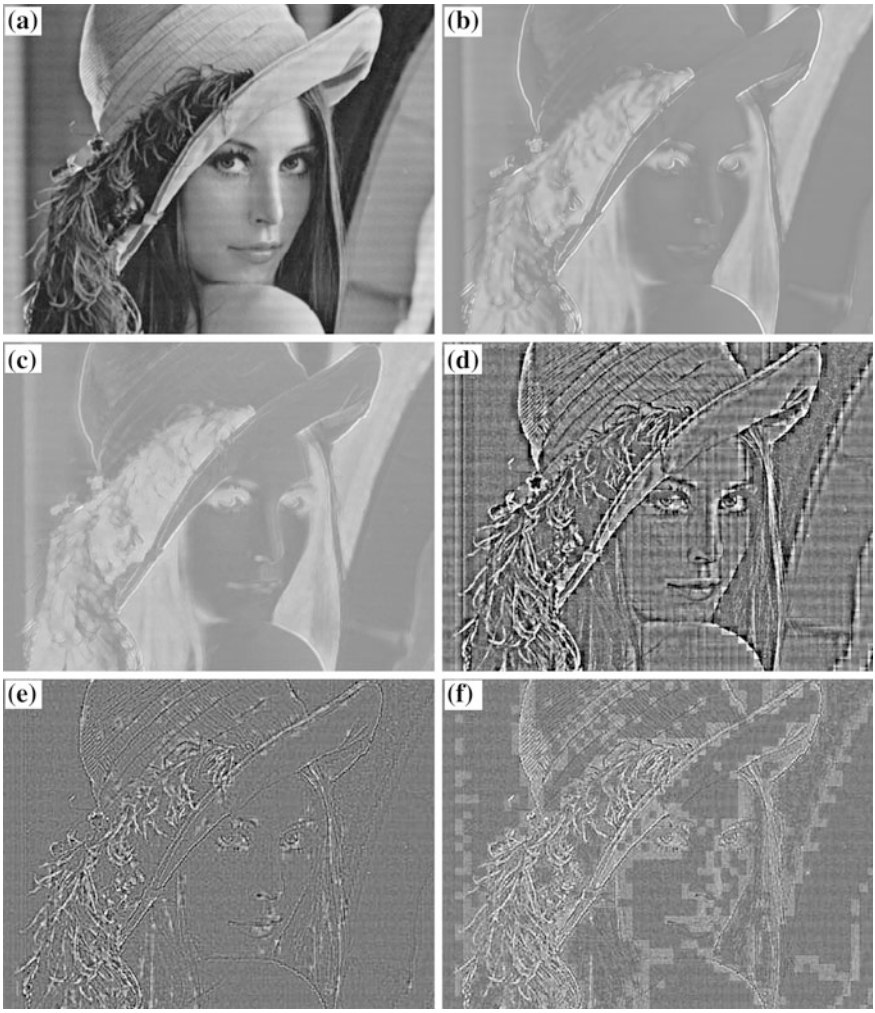

Fig. 4.7 Spatial JND profiles of TI-6 with grating at 4 cpd. Image of TI-6 with grating at 4 cpd is shown in (**a**). Spatial JND profiles obtained using Chou's, Yang's, Watson's, Wei's, and Zhang's methods are shown in (**b**–**e**), respectively

spatial JND profiles from the subband-based JND models do not reveal the horizontal gratings in TI-10.

Similar to the observations of Figs. 4.7 and 4.8, the pixel-based JND models do not detect any horizontal grating in TI-10. The observations from Figs. 4.7, 4.8 and 4.9 revealed that the pixel-based JND models are unable to detect any horizontal grating even though the grating at 4 cpd is highly visible in Fig. 4.7(a). It is clear that that the pixel-based JND models are unable to detect the modulated noise in the image due to the omission of CSF.

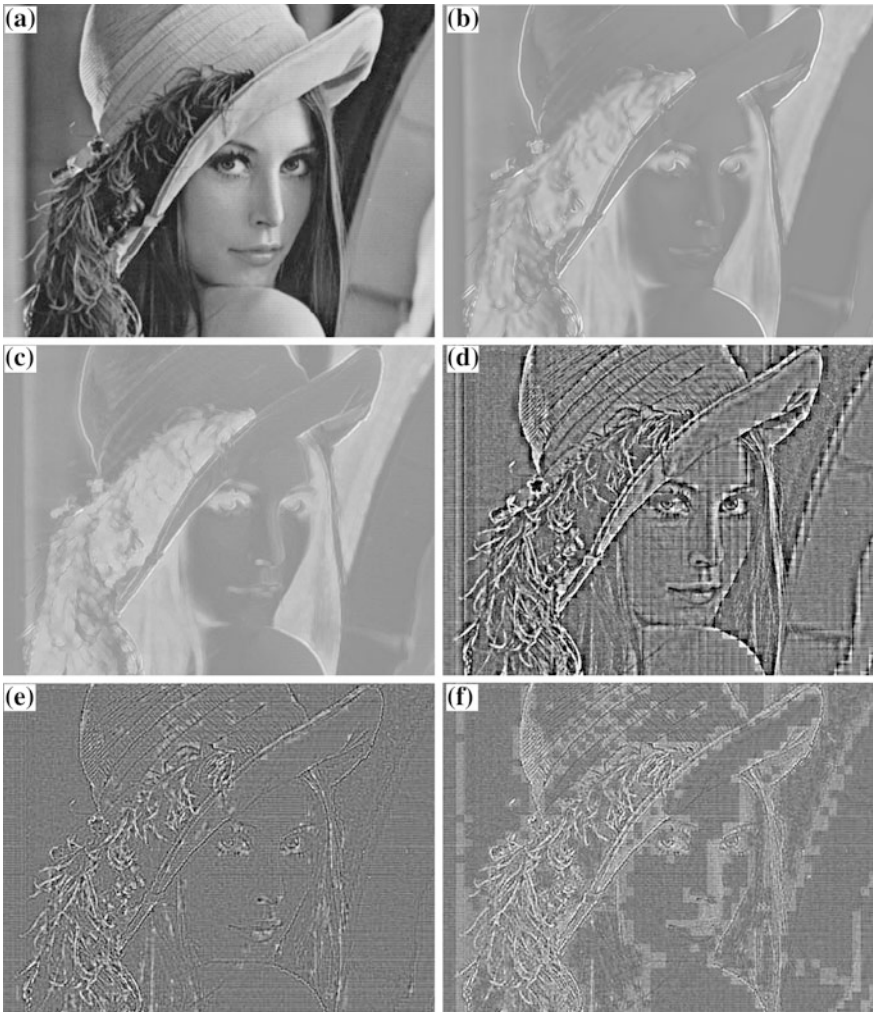

Fig. 4.8 Spatial JND profiles of TI-6 with grating at 20 cpd. Image of TI-6 with grating at 20 cpd is shown in (**a**). Spatial JND profiles obtained using Chou's, Yang's, Watson's, Wei's, and Zhang's methods are shown in (**b**–**e**), respectively

4.3 Performance Analysis

Ideally, the perceptual distortion score P computed from a JND model should correlates with the subjective score s of an image. The relation between P and s of an image can be non-linear as subjective evaluation might have non-linear compression at the extremes of the subjective score. To minimize the non-linearity of subjective evaluation, and to facilitate comparison of various JND models, the

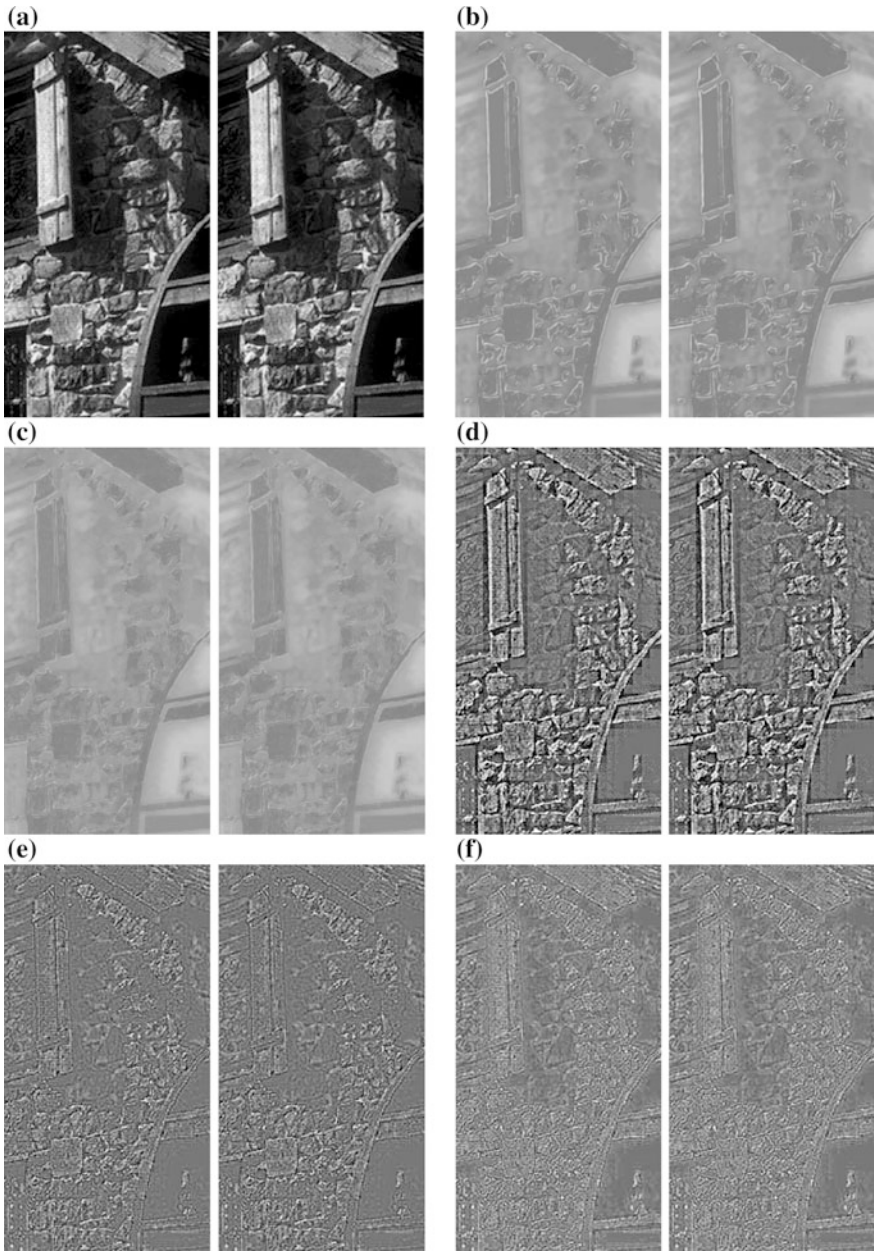

Fig. 4.9 Cropped spatial JND profiles of TI-10 with horizontal grating at 4 (*left image*) and 20 cpd (*right image*). Cropped images with horizontal gratings at 4 and 20 cpd are shown in (**a**), and spatial JND profiles obtained using Chou's, Yang's, Watson's, Wei's, and Zhang's methods are shown in (**b**–**f**), respectively

relation between a JND model's distortion score and its subjective score is estimated using non-linear regression [VQE03]. Based on [VQE03], the score $s'(P)$ is curve-fitted using a four-parameter cubic polynomial

$$s'(P) = a_0 + a_1 P + a_2 P^2 + a_3 P^3, \tag{4.4}$$

where a_0, a_1, a_2, and a_3 are fitted to a P versus s curve. VQEG employs the Pearson correlation ρ_p and Spearman rank-order correlation ρ_s to assess the prediction accuracy and prediction monotonicity of a distortion metric, respectively. In addition, the root mean square (RMS) error between $s'(P)$ and the actual subjective score s is computed to measure the accuracy of $s'(P)$. Pearson correlation and Spearman rank-order correlation are defined as follow:

$$\rho_p = \frac{\sum_k (s'(P_k) - \bar{s})(P_k - \bar{P})}{\sqrt{\sum_k (s'(P_k) - \bar{s})^2}\sqrt{\sum_k (P_k - \bar{P})^2}}, \tag{4.5}$$

and

$$\rho_s = \frac{\sum_k (\chi_k - \bar{\chi})(\gamma_k - \bar{\gamma})}{\sqrt{\sum_k (\chi_k - \bar{\chi})^2}\sqrt{\sum_k (\gamma_k - \bar{\gamma})^2}}, \tag{4.6}$$

where $s'(P_k)$ is the predicted score for test-image k and $\bar{s}$ is the mean of $s'(P)$; P_k is the perceptual distortion score of test-image k and $\bar{P}$ is the mean of P; χ_k is the rank-ordered series of $s'(P_k)$ and $\bar{\chi}$ is the mid-rank of $s'(P_k)$; γ_k is the rank-ordered series of P_k and $\bar{\gamma}$ is the mid-rank of P_k. The Pearson correlation and Spearman rank-order correlation range between -1 and 1, and $|\rho_p|$, $|\rho_s|$ become one when there is a perfect match between s' and P. RMS error s_e is defined as

$$s_e = \frac{1}{N_k}\sum_k (s'(P_k) - s_k)^2, \tag{4.7}$$

where N_k is the number of test-images and s_k is the subjective score for test-image k.

4.3.1 Comparative Analysis of PICs

In this sub-section, we compare the compression performance of three PICs based on the DCT subband-based JND models (t_{Wat}, t_{ZLX}, and t_{WN}). Some adjustments are made to these JND models before the comparative analysis is performed. First, we used the error pooling scheme given by (2.41) in Wei and Ngan's JND model [WN09] since Wei and Ngan didn't define an error pooling stage in their JND model. Second, the luminance adaptation in t_{Wat} may lead to $P(\mathbf{n})$ becoming

infinity. This problem is caused by the elevation parameter $e_1^{\text{Wat}}(\mathbf{n})$ becoming zero when the DC coefficient of a $N \times N$ DCT-II block is zero, which in turn leads to JND threshold $t(\mathbf{k}, \mathbf{n}) = 0$ and $P(\mathbf{n}) = \infty$. To prevent $P(\mathbf{n})$ becoming infinity, the elevation parameter $e_1^{\text{Wat}}(\mathbf{n})$ is modified to

$$e_1'^{\text{Wat}}(\mathbf{n}) = \begin{cases} \left(\frac{1}{C_{L_a}}\right)^{\alpha^T}, & \text{for } C(0,0,\mathbf{n}) = 0, \\ \left(\frac{C(0,0,\mathbf{n})}{C_{L_a}}\right)^{\alpha^T}, & \text{otherwise.} \end{cases} \tag{4.8}$$

To obtain compressed images at the desired perceptual distortion score, the quantization matrix used in the three PICs is computed using a scheme adapted from [ZLX05]. This scheme computes the pooled error using (2.37) for Watson's PIC and (2.41) for Wei's and Zhang's PICs. The quantization matrix $\mathbf{Q}_d$ to produce the desired perceptual score is computed using the pseudocode shown in Fig. 4.10.

Compressed images at bitrates from 0.3 to 0.5 bpp are selected for our comparative analysis of the three PICs. These compressed images are reconstructed and compared with the original images. The confidence interval of the subjective score for these compressed images is shown in Fig. 4.11, and it is found that the compressed images from the Zhang's PIC have the best subjective score. At 0.3 and 0.4 bpp, Wei's and Zhang's PICs performed similarly, and Zhang's and Watson's PICs constantly performs the best and worse among the three PICs, respectively.

Compressed images obtained with Watson's PIC tend to exhibit the most blocking artifacts among the three PICs, and the image quality of the compressed images is often degraded due to significant blocking artifacts. Interestingly, several compressed images obtained using Watson's PIC are found to have significant sharper edges and textures as compared to the compressed images from other PICs. Many compressed images obtained with Wei's PIC exhibit significant blurring artifacts, which reduce the subjective score of these images.

```
Set all elements of Q_d to one
Set k_1 = 0 and k_2 = 0
Do {
    Do {
        Increase kth element of Q_d by one
        Error pooling using ((2.36) and (2.37)) or ((2.41) and (2.42))
    } while P(k) ≤ Desired perceptual distortion score at kth subband
Go to next subband
} while k_1 < N and k_2 < N
```

Fig. 4.10 Pseudocode to obtain quantization matrix at desired perception distortion score

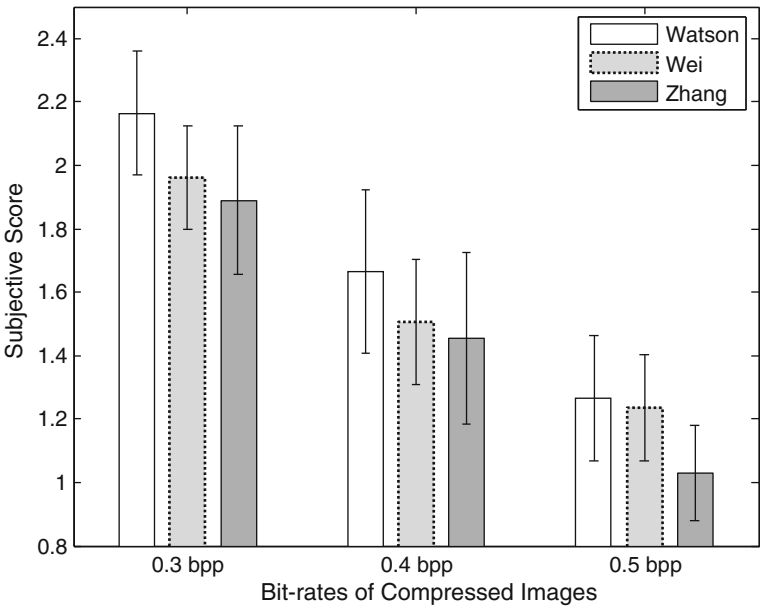

Fig. 4.11 Subjective score and 95 % confidence interval of compressed images at bitrates of 0.3–0.5 bpp

Next, we compare the correlation of the perceptual distortion score and subjective score of the three PICs. The scatterplots between the perceptual distortion score and subjective score are shown in Fig. 4.12. The scatterplot in Fig. 4.12a clearly illustrates that Watson's PIC is the least correlated with subjective score as well as having the widest spread of points. Comparably, Wei's PIC exhibits the highest correlation with subjective score as well as having the smallest spread of points.

The 95 % confidence interval of the Pearson correlation and Spearman rank-order correlation of the three PICs is shown in Fig. 4.13. As expected, Watson's PIC is found to have the largest confidence interval among the 3 PICs. The scatterplot of the Wei's PIC exhibits the highest correlation between the subjective score and perceptual distortion score, and this PIC also has the smallest confidence interval among the three PICs. As shown in Fig. 4.12, Wei's PIC has the highest Pearson correlation and Spearman rank-order correlation among the three PICs. From Figs. 4.12 and 4.13, it is observed that Zhang's and Wei's PICs perform similarly, with Wei's PIC marginally better than the Zhang's PIC. This observation coincides with those made on the noise-shaping performance of the various JND models discussed in Sect. 4.1. However, it should be noted that Zhang's PIC produces slightly better images than Wei's PIC at 0.5 bpp.

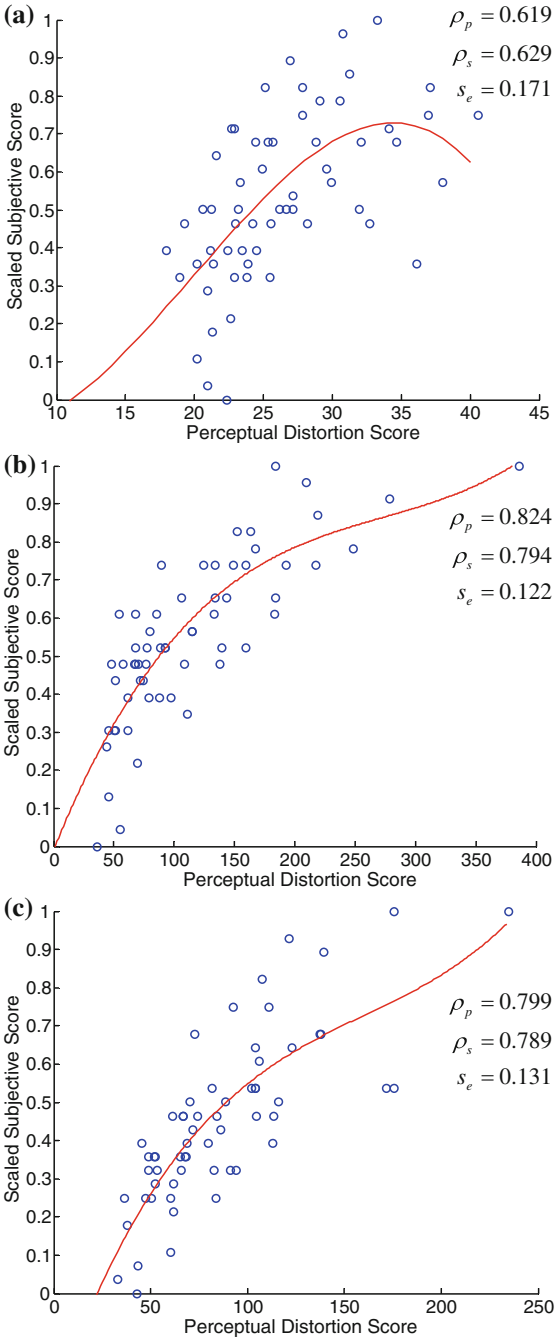

Fig. 4.12 Scatterplots of perceptual distortion score versus subjective score. Fitted polynomial function is shown in red. Scatterplots of Watson's, Wei's, and Zhang's PICs are shown in (**a**–**c**), respectively

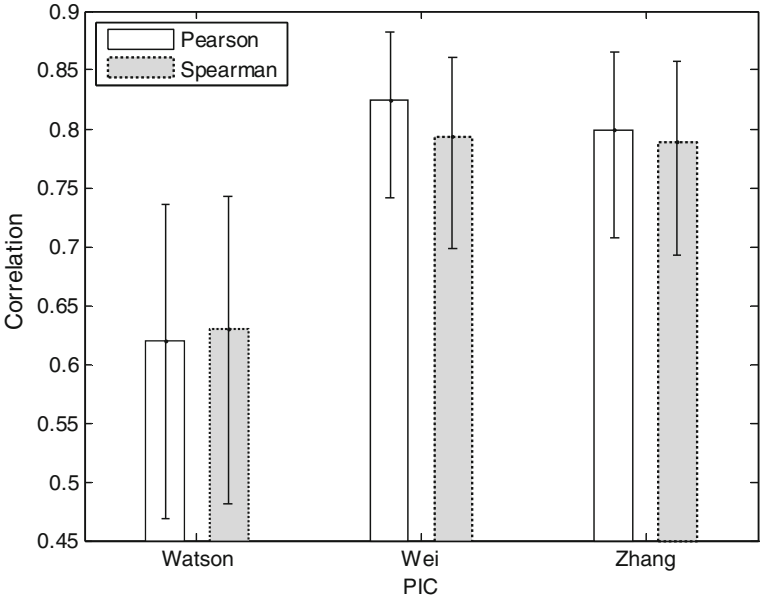

Fig. 4.13 Scatterplots of subjective score versus perceptual distortion score. Fitted polynomial function is shown in red. Scatterplots of Watson's, Wei's, and Zhang's PICs are shown in (**a**–**c**), respectively

4.4 Summary

This chapter starts with the technique to convert computational models for JND in the DCT-II domain to the pixel domain. This technique facilitates the comparison of JND models in different domains, and five JND models in the DCT-II and pixel domains are discussed in this chapter. Our comparative analysis involves Chou's and Yang's JND models in the pixel domain and Watson's, Wei's, and Zhang's JND models in the DCT-II domain. A series of subjective experiments were conducted to evaluate the performance of these JND models, and to determine their correlation with human perception.

Noise-contaminated images are generated using the spatial JND profiles from the five JND models. Our subjective experiments revealed that the spatial JND profiles computed in the DCT-II domain are generally better than those in the pixel domain, and Wei's JND model is found to have the best noise-shaping performance among the five JND models. Due to the block classification, Wei's and Zhang's JND models produce better estimation of contrast masking in the edge (lower contrast masking) and texture (higher contrast masking) regions. One distinct difference between the spatial JND profiles computed in the pixel and DCT-II domains are those estimated from the DCT-II domain tends to exhibit block-like characteristic in

the spatial JND profile, and the Watson's JND model exhibits the most abrupt changes which led to perceptible distortion.

The inferior performance of JND models in the pixel domain can be attributed to the omission of CSF. This observation is validated with our subjective experiments using noise-contaminated images with grating function (4 and 20 cpd). It is interesting to note that spatial JND profiles computed in the pixel domain are superior in discerning grain noise in images as compared to those computed in the DCT-II domain.

Error pooling is usually employed to obtain a single distortion score from the error map of a compressed image. Different error pooling schemes from Watson's, Wei's, and Zhang's JND models are compared using subjective experiments. Our subjective experiments revealed that Zhang's scheme produces the best subjective scores for compressed images at 0.3–0.5 bpp. The scheme proposed by Watson produces the most blocking artifacts which significantly degrades the visual quality of the compressed images. On the other hand, Wei's scheme produced burring artifact in many compressed images.

Based on the guidelines detailed in the report from VQEG, the Pearson correlation and Spearman rank-order correlation are used to access the prediction accuracy and prediction monotonicity of a distortion metric with respect to human perception. Using the Pearson correlation and Spearman rank-order correlation, it is found that the Wei's JND model has the highest correlation with the human perception.

Chapter 5
Concluding Remarks

This monograph has described the JND threshold for DCT-II subband, which is formulated as a product of the detection threshold and the elevation parameters, namely, luminance adaptation and contrast masking. Based on this formulation, we review Chou's [CL95], Yang's [YLL05], Watson's [Wat93], Wei's [WN09], and Zhang's [ZLX05] JND models. We have also shown how these JND models can be integrated into DCT-II based PICs, which are compatible with the JPEG standard. Although there has been significant development in the computational models for JND, a number of challenges still remain in this field. Here, we list some of the challenges as well as possible extensions of computational models for JND.

- Extensive efforts were undertaken by VQEG to develop a systematic approach to validate an objective video quality model. For the methodology defined by VQEG [VQE03], Pearson correlation and Spearman rank-order correlation are used to rank video quality models under examination. However, it should not infer that Pearson correlation and Spearman rank-order correlation are the best statistics to describe the relation between the objective score from a video quality model and the subjective score from a human viewer. Since this relation between these scores need not be linear, questions remain on the validity of applying Pearson correlation to measure the statistics of data from such a model. On this note, it may be worthwhile to investigate and determine a statistic measure that is more suited to model the nonlinearity in the relation of (if any) the scores between a video quality model and subjective viewing.
- One of the key parameters of a computational model for JND is the viewing distance of an image under examination. This monograph considered several pixel-based JND models from [CL95, YLL03, YLL05], which defined the viewing distance as six times of the image height. It would be interesting to analyze the behaviour of the JND threshold with respect to the viewing distance, and to determine the relation between the JND threshold and the viewing distance. As the computational cost of most JND models tend to be high, such a relation can be employed to provide a good estimate of JND threshold at different viewing distances.

E.-L. Tan and W.-S. Gan, *Perceptual Image Coding with Discrete Cosine Transform*, SpringerBriefs in Signal Processing,
DOI 10.1007/978-981-287-543-3_5

- With the emergence of mobile multimedia capable devices and availability of broadband networks, users have easy access to HD image and video contents. The mobile devices available to users usually have smaller displays as well as limited processing capabilities. Hence, an efficient method to reduce the spatial resolution of image and video content before delivery to these mobile devices would be useful [VCS03]. Limited storage and bandwidth have long necessitated the use of compressed image and video contents, such as JPEG, MPEG-1/2/4, and H.261/3/4. Therefore, it is highly desirable to be able to directly manipulate these contents in their compressed form. Such an approach eliminates the need to decompress and compress the image and video contents, which is necessary for pixel domain based techniques. Furthermore, it has been shown that resizing images in the DCT domain [DA01, PPO03, ST04, MM05, PP06, SC07, LL07] produces visually better images than pixel domain techniques. However, these techniques do not consider the HVS in the transcoding process. A possible extension of the JND model is perceptually-tuned image transcoding [TG12], and to establish a relation between image or video transcoding and HVS. This relation allows optimal selection of filter banks for the up-sampling and down-sampling stages in the transcoding process, which can effectively reduce ringing and blurring artifacts in the resized image and video contents.

References

[AP92] A.J. Ahumada Jr., H.A. Peterson. Luminance-model-based DCT quantization for color image compression, in *Human Vision, Visual Processing, and Digital Display III*, ed. by B.E. Rogowitz. *Proceedings of the SPIE*, 1992

[BC69] C. Blakemore, F.W. Campbell, On the existence of neurones in the human visual system selectivity to the orientation and size of retinal images. J. Physiol. **203**(1), 237–260 (1969)

[BKW75] M.A. Berkley, F. Kitterle, D.W. Watkins, Grating visibility as a function of orientation and retinal eccentricity. Vis. Res. **15**(2), 239–244 (1975)

[Bov05] A.C. Bovik, in *Handbook of Image and Video Processing* (*Communication, Networking and Multimedia*) (Academic Press, Waltham, 2005)

[Can86] J. Canny, A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-8**(6), 679–698 (1986)

[CB99] Y.J. Chiu, T. Berger, A software-only videocodec using pixelwise conditional differential replenishment and perceptual enhancement. IEEE Trans. Circuits Syst. Video Technol. **9**(3), 438–450 (1999)

[CC96] C.H. Chou, C.W. Chen, A perceptually optimized 3-D subband codec for video communication over wireless channels. IEEE Trans. Circuits Syst. Video Technol. **6**(2), 143–156 (1996)

[CL95] C.H. Chou, Y.C. Li, A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. IEEE Trans. Circuits Syst. Video Technol. **5**(6), 467–476 (1995)

[CR68] F.W. Campbell, J.G. Robson, Application of Fourier analysis to the visibility of gratings. J. Physiol. **197**(3), 551–566 (1968)

[CR90] B. Chitprasert, K.R. Rao, Human visual weighted progressive image transmission. IEEE Trans. Commun. **38**(7), 1040–1044 (1990)

[CSE00] C. Christopoulos, A. Skodras, T. Ebrahimi, The JPEG2000 still image coding system: an overview. IEEE Trans. Consum. Electron. **46**(4), 1103–1127 (2000)

[Cor90] T.N. Cornsweet, *Visual Perception* (Academic Press, Waltham, 1990)

[DA01] R. Dugad, N. Ahuja, A fast scheme for image size change in the compressed domain. IEEE Trans. Circuits Syst. Video Technol. **11**(4), 461–474 (2001)

[Dal92] S. Daly, Visible difference predictor: an algorithm for the assessment of image fidelity, in *Proceedings of the SPIE*, vol. 2 (1992), pp. 2–15

[Dal93] S. Daly, The visible difference predictor: an algorithm for the assessment of image fidelity, in *Digital Images and Human Vision* (1993), pp. 179–206

[Dau80] J.G. Daugman, Two-dimensional spectral analysis of cortical receptive field profiles. Vis. Res. **20**(10), 847–856 (1980)

[DAT82] R.L. DeValois, D.G. Albrecht, L.G. Thorell, Spatial frequency selectivity of cells in the macaque visual cortex. Vis. Res. **22**(5), 545–559 (1982)

E.-L. Tan and W.-S. Gan, *Perceptual Image Coding with Discrete Cosine Transform*, SpringerBriefs in Signal Processing,
DOI 10.1007/978-981-287-543-3

[DYH82] R.L. DeValois, E.W. Yund, N. Hepler, The orientation and direction selectivity of cells in macaque visual cortex. Vis. Res. **22**(5), 531–544 (1982)

[EB98] M.P. Eckert, A.P. Bradley, Perceptual quality metrics applied to still image compression. Signal Process. **70**(3), 177–200 (1998)

[Fau79] O.D. Faugeras, Digital color image processing within the framework of a human visual model. IEEE Trans. Acoust. Speech Signal Process. **27**(4), 380–393 (1979)

[Gir93] B. Girod, What's wrong with mean-squared error, in *Digital Images and Human Vision*, MIT Press, pp. 207–220, 1993.

[GRN78] N. Graham, J.G. Robson, J. Nachmias, Grating summation in fovea and periphery. Vis. Res. **18**(7), 815–825 (1978)

[Hec24] S. Hecht, The visual discrimination of intensity and the Weber-Fechner law. J. Gen. Physiol. **7**(2), 235–267 (1924)

[HL83] H.C. Reeve, J.S. Lim, Reduction of blocking effect in image coding. IEEE Int. Conf. Acoust. Speech Signal Process. **8**, 1212–1215 (1983)

[HK00] I. Höntsch, L.J. Karam, Locally adaptive perceptual image coding. IEEE Trans. Image Process. **9**(9), 1472–1483 (2000)

[HK02] I. Höntsch, L.J. Karam, Adaptive image coding with perceptual distortion control. IEEE Trans. Image Process. **11**(3), 213–222 (2002)

[HW62] D.H. Hubel, T.N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. **160**(1), 106–154 (1962)

[ISO94] *Information Technology—Digital Compression and Coding of Continuous-Tone Still Images: Requirements and Guidelines*, ISO/IEC 10918-1, 1994

[ISO00] *Information Technology—JPEG 2000 Image Coding System—Part 1: Core Coding System,* ISO/IEC 15444-1, 2000

[ITU02] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R BT.500-11, 2002

[JJS93] N. Jayant, J. Johnston, R. Safranek, Signal compression based on models on human perception. Proc. IEEE **81**(10), 1385–1422 (1993)

[Joh80] J.D. Johnston, A filter family designed for use in quadrature mirror filter banks. IEEE Int. Conf. Acoust. Speech Signal Process. **5**, 291–294 (1980)

[Kel85] D.H. Kelly, Visual processing of moving stimuli. J. Opt. Soc. Am. **2**(2), 216–225 (1985)

[Leg78a] G.E. Legge, Sustained and transient mechanisms in human vision: temporal and spatial properties. Vis. Res. **18**(1), 69–81 (1978)

[Leg78b] G.E. Legge, Space domain properties of a spatial frequency channel in human vision. Vis. Res. **18**(8), 959–969 (1978)

[LF80] G.E. Legge, J.M. Foley, Contrast masking in human vision. J. Opt. Soc. Am. **70**(12), 1458–1471 (1980)

[LK00] Y.K. Lai, C.C.J. Kuo, A Haar wavelet approach to compressed image quality measurement. J. Vis. Commun. Image Represent. **11**, 17–40 (2000)

[LL07] Y.R. Lee, C.W. Lin, Visual quality enhancement in DCT-domain spatial downscaling transcoding using generalized DCT decimation. IEEE Trans. Circuits Syst. Video Technol. **17**(8), 1079–1084 (2007)

[LLP10] A.M. Liu, W.S. Lin, M. Paul, C.W. Deng, F. Zhang, Just noticeable difference for images with decomposition model for separating edge and textured regions. IEEE Trans. Circuits Syst. Video Technol. **20**(11), 1648–1652 (2010)

[LMK98] B. Li, G.W. Meyer, R.V. Klassen, A comparison of two image quality models. SPIE Conf. Hum. Vis. Electron. Imaging III **3299**, 98–109 (1998)

[Loh84] H. Lohscheller, A subjective adapted image communication system. IEEE Trans. Comm. **COM-32**(12), 1316–1322 (1984)

[Lub93] J. Lubin, The use of psychophysical data and models in the analysis of display system performance, in *Digital Image and Human Vision* (1993), pp. 163–178

[Lub95] J. Lubin, A visual discrimination mode for image system design and evaluation, *Visual Models for Target Detection and Recognition* (1995), pp. 207–220

[MM05] J. Mukherjee, S.K. Mitra, Arbitrary resizing of images in DCT space. IEE Proc. Vis. Image Signal Process. **152**(2), 155–164 (2005)

[MSO07] R. Muralishankar, A. Sangwan, D. O'Shaughnessy, Theoretical complex cepstrum of DCT and warped DCT filters. IEEE Signal Process. Lett. **14**(5), 367–370 (2007)

[MS74] J.L. Mannos, D.J. Sakrison, The effects of a visual fidelity criterion on the encoding of images. IEEE Trans. Inform. Theor. **20**(4), 525–536 (1974)

[MTT78] J.A. Movshon, I.D. Thompson, D.J. Tolhurst, Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. J. Physiol. **283**, 101–120 (1978)

[NAW93] R. Navarro, P. Artal, D.R. Williams, Modulation transfer of the human eye as a function of retinal eccentricity. J. Opt. Soc. Am. **10**(2), 201–212 (1993)

[NB67] F.L. van Nes, M.A. Bouman, Spatial modulation transfer in the human eye. J. Opt. Soc. Am. **57**(3), 401–406 (1967)

[NLS86] K.N. Ngan, K.S. Leong, H. Singh, Cosine transform coding incorporating human visual system model, in *SPIE Fiber* (1986), pp. 165–171

[Nil85] N.B. Nill, A visual model weighted cosine transform for image compression and quality assessment. IEEE Trans. Commun. **33**(6), 551–557 (1985)

[PAW93a] H.A. Peterson, A.J. Ahumada, A.B. Watson, An improved detection model for DCT coefficient quantization, in *Proceedings of the SPIE*, vol. 1913 (1992), pp. 191–201

[PAW93b] H.A. Peterson, A.J. Ahumada, A.B. Watson, The visibility of DCT quantization noise, in *Society for Information Display Digest of Technical Papers* (1993), pp. 942–945

[PDT77] G.F. Poggio, R.W. Doty, W.H. Talbot, Foveal striate cortex of behaving monkey: single-neuron responses to square-wave gratings fixation of gaze. J. Neurophysiol. **40**(6), 1369–1391 (1977)

[PJJ94] J. Park, J.M. Jo, J. Jeong, Some adaptive quantizers for HDTV image compression. Signal Process. HDTV (1994)

[PMP91] H.A. Peterson, H. Peng, J.H. Morgan, W.B. Pennebaker, Quantization of color image components in the DCT domain, in *Human Vision, Visual Process. Digital Display II*, *Proceeding of SPIE*, vol. 1453 (1991), pp. 210–222

[PP06] Y.S. Park, H.W. Park, Arbitrary-ratio image resizing using fast DCT of composite length for DCT-based transcoder. IEEE Trans. Image Process. **15**(2), 494–500 (2006)

[PPO03] H.W. Park, Y.S. Park, S.K. Oh, L/M-fold image resizing in block-DCT domain using symmetric convolution. IEEE Trans. Image Process. **12**(9), 1016–1034 (2003)

[PW84] G.C. Philips, H.R. Wilson, Orientation bandwidths of spatial mechanisms measured by masking. J. Opt. Soc. Am. **1**(2), 226–232 (1984)

[RAW97] A.M. Rohaly, A.J. Ahumada, A.B. Watson, Object detection in natural backgrounds predicted by discrimination performance and models. Vis. Res. **37**(23), 3225–3235 (1997)

[RG81] J.G. Robson, N. Graham, Probability summation and regional variation in contrast sensitivity across the visual field. Vis. Res. **21**(3), 409–418 (1981)

[SC07] H. Shu, L.P. Chau, A resizing algorithm with two-stage realization for DCT-based transcoding. IEEE Trans. Circ. Syst. Video Technol. **17**(2), 248–253 (2007)

[Sch56] O.H. Schade, Optical and photoelectric analog of the eye. J. Opt. Soc. Am. **46**(9), 721–739 (1956)

[SE00] D. Santa-Cruz, T. Ebrahimi, A study of JPEG 2000 still image coding versus other standards, in *Proceedings of X European Signal Processing Conference*, vol. 2 (2000), pp. 673–676

[SJ89] R.J. Safranek, J.D. Johnston, A perceptually tuned subband image coder with image dependent quantization and post-quantization, in *IEEE International Conference on Acoustic, Speech, Signal Processing* (1989), pp. 1945–1948

[SJ72] C.F. Stromeyer, B. Julesz, Spatial-frequency masking in vision: critical bands and spread of masking. J. Opt. Soc. Am. **62**(10), 1221–1232 (1972)

[ST04] C.L. Salazar, T.D. Tran, On resizing images in the DCT domain. IEEE Int. Conf. Image Process. **4**, 2797–2800 (2004)

[SW96] D. Shen, S. Wang, Measurements of JND property of HVS and its applications to image segmentation, coding and requantization, in *Proceedings of the SPIE Digital Computing Technical System Video Communication* (1996), pp. 113–121

[TG11] E.L. Tan, W.S. Gan, Perceptually tuned subband coder for JPEG. J. Real Time Image Process. **6**(2), 101–115 (2011)

[TG12] E.L. Tan, W.S. Gan, Perceptual image coding and transcoding with subband discrete cosine transform, Ph.D. Dissertation, Nanyang Technological University, 2012

[TH94a] P.C. Teo, D.J. Heeger, Perceptual image distortion, in *Proceedings of the SPIE*, vol. 2179 (1994), pp. 127–141

[TH94b] P.C. Teo, D.J. Heeger, Perceptual image distortion, in *Proceedings of the IEEE International Conference on Image Process*ing, vol. 2 (1994), pp. 982–986

[TS96] T.D Tran, R. Safranek, A locally perceptual masking threshold model for image coding, in *IEEE International Conference on Acoustic, Speech, Signal Processing* (1996), pp. 1882–1885

[TV98] H.Y. Tong, A.N. Venetsanopoulos, A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking, in *Proceedings of the EEE International Conference on Image Processing* (1998)

[URL01] http://decsai.ugr.es/cvg/dbimagenes/g512.php

[VCS03] A. Vetro, C. Christopoulos, H. Sun, Video transcoding architectures and techniques: an overview. IEEE Signal Process. Mag. **20**(2), 18–29 (2003)

[VQE03] VQEG, *Final Report from Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment* (2003). Available ftp://vqeg.its.bldrdoc.gov/Documents/VQEG_Approved_Final_Reports/VQEGII_Final_Report.pdf

[WA05] A.B. Watson, A.J. Ahumada, A standard model for foveal detection of spatial contrast. J. Vis. **5**(9), 717–740 (2005)

[Wal92] G.K. Wallace, The JPEG still picture compression standard. IEEE Trans. Consum. Electron. **38**(1), xviii–xxxiv (1992)

[Wat79] A.B. Watson, Probability summation over time. Vis. Res. **19**(5), 515–522 (1979)

[Wat82] A.B. Watson, Summation of grating patches indicates many types of detectors at one retinal location. Vis. Res. **22**(1), 17–25 (1982)

[Wat93] A.B. Watson, DCT quantization matrices visually optimized for individual images, in *Proceedings of the SPIE Human Vision, Visual Proceedings, Digital Display IV* (1993), pp. 202–216

[WHM97] A.B. Watson, G.Y. Yang, J.A. Solomon, J. Villasenor, Visibility of wavelet quantization noise. IEEE Trans. Image Process. **6**(8), 1164–1175 (1997)

[WLM90] H.R. Wilson, D. Levi, L. Maffei, J. Rovamo, R. DeValois, The perception of form: Retina to striate cortex, in *Visual Perception: The Neurophysiological Foundations*, ed. by L. Spillman, J. Werner (Academic Press, Waltham, 1990), pp. 231–272

[WN09] Z.Y. Wei, K.N. Ngan, Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain. IEEE Trans. Circ. Syst. Video Technol. **19**(3), 337–346 (2009)

[WR84] H.R. Wilson, D. Regan, Spatial frequency adaptation and grating discrimination: prediction of a line element model. J. Opt. Soc. Am. **1**(11), 1091–1096 (1984)

[WSS00] M.J. Weinberger, G. Seroussi, G. Sapiro, The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS. IEEE Trans. Image Process. **9**(8), 1309–1324 (2000)

[YLL03] X.K. Yang, W.S. Lin, Z.K. Lu, E.P. Ong, S.S. Yao, On incorporating just-noticeable-distortion profile into motion-compensated prediction for video compression. IEEE Int. Conf. Image Process. **3**, 833–836 (2003)

[YLL05] X.K. Yang, W.S. Lin, Z.K. Lu, E.P. Ong, S.S. Yao, Just noticeable distortion model and its applications in video coding. Signal Process. Image Commun. **20**(7), 662–680 (2005)

[ZLX05] X.H. Zhang, W.S. Lin, P. Xue, Improved estimation for just-noticeable visual distortion. Signal Process. **85**(4), 795–808 (2005)

[ZLX08] X.H. Zhang, W.S. Lin, P. Xue, Just-noticeable-difference estimation with pixels in images. J. Vis. Commun. Image Represent. **19**(1), 30–41 (2008)